The power of 1

Interventions and outcome measures for rare genetic neurodevelopmental disorders

Annelieke Rosalie Müller

The power of 1

Interventions and outcome measures for rare genetic neurodevelopmental disorders

Annelieke Rosalie Müller

This research was funded by Stichting Wetenschappelijk Onderzoek 's Heeren Loo

Cover design: Jasper Müller
Lay out: Ferdinand van Nispen, Citroenvlinder DTP & Vormgeving, *my*-thesis.nl, Ede
Print: Proefschriftenprinten.nl, Ede, The Netherlands

ISBN: 978-90-833743-7-6

Please cite as: Müller, A.R. (2024). The power of 1: Interventions and outcome measures for rare genetic neurodevelopmental disorders. Amsterdam UMC location University of Amsterdam, Department of Pediatric Metabolic Diseases, Amsterdam, the Netherlands.

The power of 1

Interventions and outcome measures for rare genetic neurodevelopmental disorders

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. P.P.C.C. Verbeek
ten overstaan van een door het College voor Promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op vrijdag 26 januari 2024, te 13.00 uur

door Annelieke Rosalie Müller
geboren te Amsterdam

Promotiecommissie

Promotor:	prof. dr. F.A. Wijburg	AMC-UvA
Copromotores:	dr. A.M. van Eeghen	AMC-UvA
	dr. E. Boot	‘s Heeren Loo
Overige leden:	prof. dr. H. Bruining	AMC-UvA
	prof. dr. L. Henneman	Vrije Universiteit Amsterdam
	prof. dr. J.W. Groothoff	AMC-UvA
	prof. dr. N.I. Wolf	Vrije Universiteit Amsterdam
	prof. dr. M. Offringa	AMC-UvA
	dr. S.A. Huisman	AMC-UvA

Faculteit der Geneeskunde

Table of contents

Chapter 1

General introduction

Millions of people worldwide are affected by one of the nearly 8,000 rare disorders, defined as a condition affecting less than one in 2,000 individuals.[1] Around 80 percent of these rare disorders has a genetic origin and many are associated with neurodevelopmental disorders.[2] These (rare) genetic neurodevelopmental disorders (RGND) are collectively common, affecting 1-3% of the population directly and another 5% indirectly, i.e., family members coping with care.[3,4] There is a marked overlap between individuals with an RGND and with an intellectual disability (ID) (Figure 1). ID is characterized by substantial limitations in both intellectual functioning and adaptive behavior, originating in childhood or adolescence.[5–7] Genetic factors play a major role in the etiology of ID, with more than 1700 ID-related genetic disorders already identified,[8] apart from exogenous factors such as birth complications, trauma, extreme malnutrition, and infection.[9]

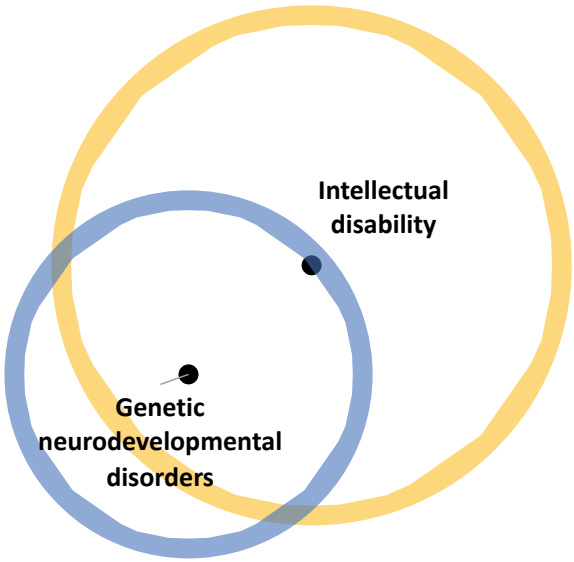


Figure 1. Schematic representation of the populations with intellectual disability and rare genetic neurodevelopmental disorders. Importantly, the figure depicts an indication rather than a precise scaled proportion.

RGNDs and ID are often accompanied by complex somatic and neuropsychiatric comorbidities, including epilepsy, sensory deficits, movement disorders, autism spectrum disorders, anxiety disorders, and cognitive and behavioral disturbances, such as irritability, aggression, and

self-injurious behavior. These comorbidities are often refractory to standard pharmacological and psychological interventions, and necessitate intensive care. This poses an enormous burden on affected individuals, families, caregivers, and healthcare systems, implying clinical and economic burden as well.[10] Because of the lifelong care needs on multiple life domains, these patient populations present significant challenges for care providers when it comes to delivering optimal personalized care. It requires specialized knowledge and expertise from various healthcare professionals, such as physicians, behavioral therapists and psychologists, occupational and speech therapists, dieticians, and direct support staff. RGNDs may involve multiple organ systems, comorbidities and unique treatment approaches. It may impact all life domains covered by the International Classification of Functioning and Disability (ICF) framework of the World Health Organization.[11]

Increasing knowledge about somatic and mental health manifestations in RGNDs and ID has resulted in improved (disorder-specific) preventive interventions and care.[12,13] For example, life expectancy has increased from ten years in 1960 to 47 years in 2007 in individuals with Down syndrome due to preventative screening and improved treatments.[14] Moreover, parkinsonism has increasingly recognized in specific RGNDs with implications for management.[15] Also, multisystem manifestations of Tuberous Sclerosis Complex (TSC) vary across age.[16] Specifically, new subependymal giant cell astrocytomas (SEGAs), which are characteristic TSC brain lesions, arise frequently during childhood, but very rarely after 25 years of age, for which frequency of performing brain imaging differs between childhood and adulthood.[17] Another example includes the often devastating and self-injurious behavior with higher prevalence in several RGNDs compared to idiopathic ID.[18] It has also been found that once pathological gastro-esophageal disease in Cornelia de Lange syndrome has been treated, behavioral symptoms significantly improve.[19,20]

Upcoming interventions and treatment options

As treatment targets are increasingly identified for many RGNDs[21,22] and interventional research has gained emerging interest[23,24], there is a great need for evidence-based therapeutic interventions at several levels

(Figure 2). Personalized treatments vary from vitamins, diets, (repurposed) drugs, organ or stem cell transplants or therapy, to enzyme, RNA and gene therapy.[25] Although often used interchangeably with precision medicine, personalized medicine refers to an N-of-1 approach that tailors interventions to the specific characteristics of an individual whereas precision medicine might be rather in line with an 1-of-N approach, using for example information from genomics, transcriptomics, metabolomics or proteomics to classify an individual into a subgroup. Personalized interventions include both targeted therapies when the intervention acts upon the underlying disease mechanism, and 'untargeted', taking into account the unique characteristics and needs of the affected individual.

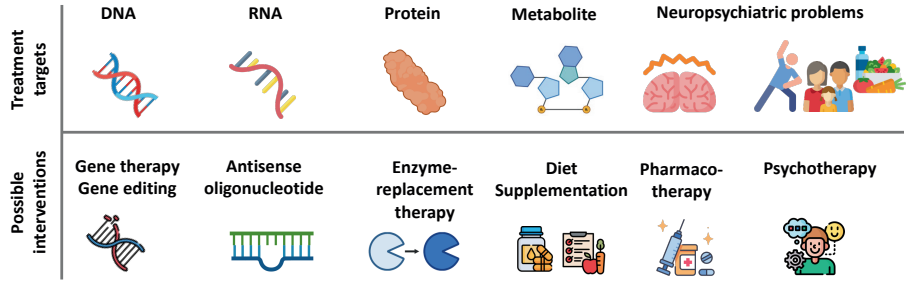


Figure 2. Level of treatment targets (upper part) and examples of corresponding possible interventions (lower part) for brain manifestations in rare genetic neurodevelopmental disorders.

Usually, multiple interventions focusing on both physical and mental health manifestations are applied in individuals with RGNDs and ID, throughout the entire lifespan. The neuropsychiatric manifestations are typically the greatest burden to the affected individuals and family.[26–28] Interventions may include both pharmacological and behavioral interventions, with risk for both over- and undertreatment.[29,30] Several targeted therapies and disease-modifying treatments are available with many more underway, such as mammalian target of rapamycin (mTOR) inhibitors for TSC or gene and enzyme therapies for metabolic disorders, including lysosomal storage disorders.[31–35] Evidence for interventions is of great importance, especially for neuropsychiatric manifestations considering its burden and adverse effects of pharmacotherapy.[26–28] Manifestations and effectiveness of interventions can differ between RGNDs, such as medication being specifically effective

for particular disorders including vigabatrin as a treatment for epilepsy in Tuberous Sclerosis Complex.[36,37] Similarly, it implies targeted behavioral interventions, such as anticipatory care planning by behavioral experts and caregivers for individuals with Down syndrome who eventually all show neuropathological changes of Alzheimer's disease at an early age.[38] Interventional research has generally performed per disorder.[39] However, effectiveness of interventions has largely remained unclear, resulting in affected individuals missing out on possibly effective interventions.

Trial methodology: how to do interventional research in RGNDs?

Interventional research is challenging in individuals with RGNDs and ID due to the rarity of the conditions and heterogeneity of manifestations. These manifestations typically show great inter- and intraindividual variability, hampering conventional trial designs. Parallel group randomized controlled trials (RCTs), considered the gold standard for interventional studies, are often not feasible in small and heterogeneous populations. Due to strict eligibility criteria, affected individuals are often excluded. To stimulate the search for treatments of these largely ignored rare disorders, a new methodological framework needs to be developed.

Single-case experimental designs

Single-case experimental designs (SCEDs) may provide a powerful solution for interventional research in RGNDs and ID, and a much-needed bridge between practice and science. SCEDs are experimental designs to test the effectiveness of an intervention for an individual participant who acts as their own control, using repeated outcome measurements and sequential or randomized introduction of the intervention.[40] SCEDs may increase acceptance by clients and physicians and may facilitate recruitment. Also, SCEDs require less participants than conventional trial designs, efficiently using data from participants.[41] Commonly used SCEDs include the multiple baseline design, the changing criterion design, the alternating treatments design, and the N-of-1 design.[42]

N-of-1 trials

Of the SCEDs, the N-of-1 design provides the highest level of evidence, as it closely follows indications of causality.[43] N-of-1 studies are randomized,

controlled, multiple crossover trials in a single patient and can provide rigorous evidence of treatment effects for individuals (Figure 3).[44,45] They address the question of inter-individual variability in treatment response, and the lack of knowledge about effects in affected individuals who are typically excluded in RCTs. Where RCTs generally assess effectiveness of treatments using the average treatment effect, N-of-1 studies can identify individual particular characteristics that may modify response to the intervention.[46] N-of-1 studies provide an ideal tool for perceiving little but significant changes and patterns over time. When the same N-of-1 design is used for several individuals, aggregated data can produce treatment effect estimates at population level which may be as robust as traditional RCTs.[47] Furthermore, the personalized approach has the potential of maximizing treatment adherence that is both patient-centered and evidence-based.[44,48–50] However, N-of-1 trials have been criticized because of challenged generalizability due to methodological and statistical bottlenecks with limited acceptance by investigators, medical ethical committees, health care institutes, and health insurance organizations.[42,51–54] Given the potential, N-of-1 studies appear to be underused, despite the urgent call for personalized medicine and the challenge the field of RGNDs are facing with regard to evidence-based medicine.

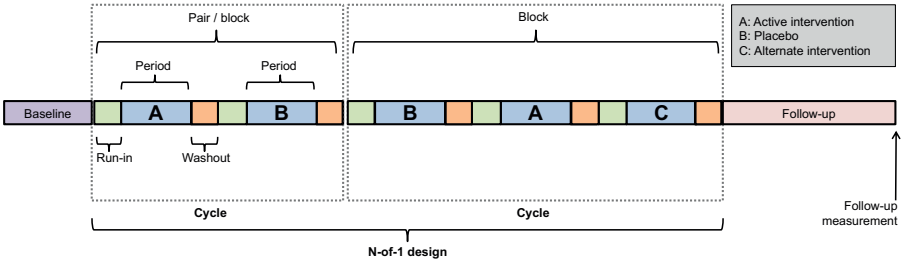


Figure 3. Schematic representation of the N-of-1 design.

Outcomes: how to measure what matters?

Due to the complexity and variety of manifestations, various outcomes have been measured using several outcome measurement instruments to assess disease severity and functioning in interventional studies. Deciding upon the right outcome measure has far-reaching implications. Sample size calculations are often based on the primary outcome measure. Trials

that do not demonstrate significant benefits based on these primary endpoints are deemed 'negative', with implications for registration and reimbursement by regulatory authorities.[55] For example, within the Fragile X syndrome community, promising targeted treatments did not show positive effects on the primary behavioral outcome measures, despite improvements on secondary outcome measures or in post hoc analyses of potentially meaningful clinical subgroups, resulting in discontinuation of drug development and clinical trials by pharmaceutical companies.[56,57] Subsequently, it was questioned whether it was the intervention that was not effective or the chosen outcome measures were not appropriate.[55,58] These negative outcomes have been disheartening to families, resulting in a call to develop well-designed studies with sensitive outcome measures.

Other hurdles have been encountered to choose appropriate outcome measures for clinical trials in ID, such as varying cognitive abilities of study participants, unavailability of outcome measures for self-report or adults with ID, limited sensitivity and lack of consensus on the best measures for this population.[55,59] Choosing appropriate specific outcome measures for clinical trials can thus be challenging, considering feasibility, utility, acceptability, and measurement properties including validity and responsiveness to change. Surrogate outcomes are used as a substitute for a direct measure of how an individual feels or functions.[60] However, surrogate or clinical outcomes are often narrow in their focus and it might be unclear whether changes are relevant. To get a better understanding of what is relevant to measure, it is important to monitor symptoms and measure the impact of the disease on functioning or quality of life. The ICF provides a framework for choosing outcome measures, capturing all components for better understanding of the disease's impact on different life domains of an individual (Figure 4).[61]

Patient-reported outcome measures

Information about functioning can be obtained from patient-reported outcome measures (PROMs).[62–64] PROMs are instruments that measure how affected individuals experience their own health and enable quantification and evaluation of severity and the impact of a disease from the patient's perspective. It can be used for both monitoring and informing care, and as an outcome measure in clinical trials.[65]

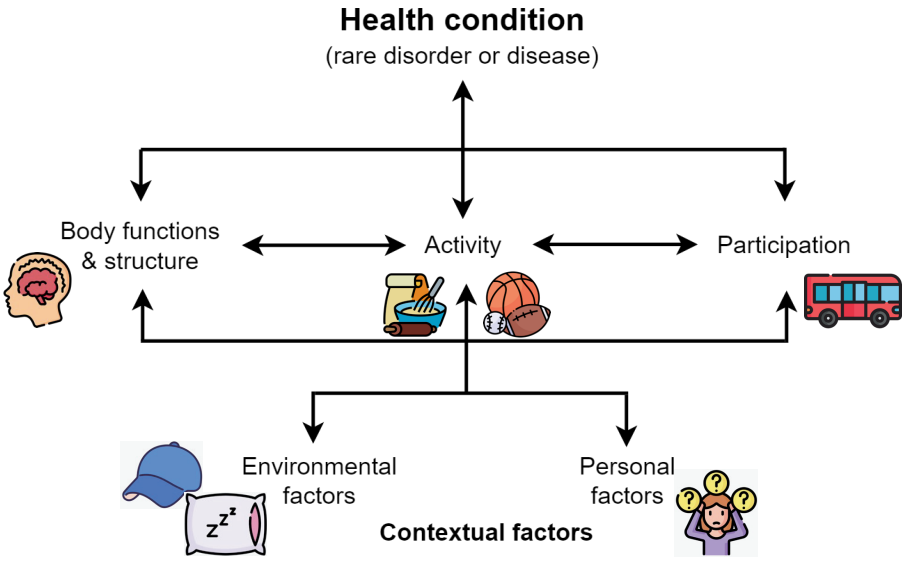


Figure 4. Impact of a rare genetic neurodevelopmental disorder on multiple life domains with different treatment targets and choice of outcome measures, based on the International Classification of Functioning and Disability (ICF) framework (World Health Organization).

However, in RGNDs, the choice of outcomes is often more complicated compared to more common diseases due to the small number of patients and the heterogeneity of the patient populations. PROMs commonly used in clinical trials do often not include disease-specific symptoms, may not be responsive enough for individuals with ID, and are often solely available as self-report.[66,67] Individuals with ID are often not able to report on their condition, and parents and caregivers are asked to complete assessments. Therefore, proxy-friendly measurement instruments, such as mobile apps, are desirable to ensure trial adherence.[68] Furthermore, threshold or ceiling effects may be more common when used in these heterogeneous populations, which limits the already small number of patients who are eligible to participate. Additionally, (adult) proxy-report instruments are often not available for the domains of interest or are too broad in scope to be sensitive enough for specific changes in particular disorders, which eventually might limit treatment adherence and acceptability as well. As such, it is suggested that health problems have been underestimated because of excluding more severely affected individuals.[69–71]

Personalized outcome measures

To measure what matters to affected individuals, their perspectives should be included, as trials are ultimately aimed at improving the affected individuals' well-being.[72] Personalized outcome measures could be used, such as Goal Attainment Scaling (GAS). GAS is an instrument that is intended for standardized evaluation of the effect of an intervention based on individualized goals.[73] In this way, quantitative expression of meaningful subjective patient experiences is enabled while translating it into evidence. However, GAS has not yet been validated for this population, neither have other personalized outcomes measures been commonly used in clinical trials.

Due to the huge amount of available outcome measurement instruments, yet without clear consensus about applicability to RGNDs and ID with regard to availability and validity, multiple tools have been selected in clinical trials, which might hamper extrapolation of trial results and acceptance by affected individuals, family and clinicians.

How to realize disorder-specific care? From diagnosis towards care

With advances in genetic and omics techniques, a genetic underlying cause can be identified in up to 50% of individuals with ID, with many more awaiting diagnosis.[74,75] As a result, more children with RGNDs or inherited metabolic disorders can be diagnosed at an early age.[76] Early diagnoses ideally allow targeted therapy to exert its effect in the crucial neurodevelopmental time window, and prevent diagnostic odysseys. Potentially, it might prevent (progression of) devastating health manifestations.[77–79]

Knowledge about the genetic etiology and associated somatic and behavioral phenotypes may provide detailed information on the cause, prognosis, inheritance for family planning, and treatment options.[20,80] It might help to understand and accept the disease by affected individuals, family, and care providers, which enables adequate and timely responding and might increase empowerment. It provides supportive care, special education or tools, access to expertise centers and peer support groups, and financial and emotional support. RGNDs are often associated with several comorbidities, as they encompass known susceptibility loci for specific comorbidities,

including vascular, hematological, endocrine diseases, and movement and eye disorders.[81,82] Screening of common comorbidities associated with specific RGNDs thus contributes to preventive care and allows for personalized medicine. Furthermore, knowing the genetic etiology reveals important information about treatments, since some may have different requirements, efficacy, and adverse events.[83] Next to advances in medical care, a diagnosis enables disorder-specific care by behavioral experts and caregivers and can improve targeted neuropsychological examinations, psychoeducation, behavioral interventions, and anticipation on supportive care, such as housing facilities.

Questions addressed in this thesis

Interventional research for individuals with complex, rare and heterogeneous disorders faces challenges with regard to trial methodology, outcome measures, and licensing and reimbursement, leading to patients missing out on possibly effective interventions. Personalized medicine has gained emerging interest, especially now underlying pathophysiological mechanisms and targeted treatments are increasingly identified and developed. Methodological quality, an N-of-1 framework, generalizability, feasibility and personalization are of paramount importance to realize the sorely needed evidence-based interventions for these vulnerable patients, enabling personalized medicine in both research and clinical practice. In this thesis, three main questions are addressed:

1. How should N-of-1 studies be performed in RGNDs?
2. How should outcome measures be used for individuals with RGNDs and ID?
3. To what extent have genetic diagnoses been reported in multidisciplinary ID care to enable disorder-specific care?

Thesis outline

This thesis is divided in three parts. **Part I** (Chapters 2 to 4) focuses on N-of-1 studies in RGNDs. **Chapter 2** describes the current literature on N-of-1 trials in RGNDs and provides recommendations for future N-of-1 studies to ultimately optimize evidence-based and personalized care.

Chapter 3 illustrates the use of the N-of-1 methodology in a protocol to investigate the effectiveness of methylphenidate for attention-deficit/hyperactivity disorder in children and adults with Smith-Magenis syndrome. In **Chapter 4** a protocol of another N-of-1 series is described in which the effectiveness of cannabidiol on complex behavior is investigated in children and adults with Tuberous Sclerosis Complex, Fragile X syndrome, and mucopolysaccharidosis type III.

Part II of this thesis focuses on outcome measures for RGNDs and ID with unknown etiology. In **Chapter 5** an overview is provided of outcomes and outcome measurement instruments used in clinical trials in RGNDs and ID, and exposes the problem with regard to reporting outcomes and available outcome measurement instruments for this population. In **Chapter 6** a newly developed and validated disorder-specific outcome measure for Tuberous Sclerosis Complex is described, called the TSC-PROM, which measures the impact of the disorder on all relevant domains of functioning. **Chapter 7** describes an accessibility and feasibility study of experience sampling methods for assessing mental health of individuals with ID, based on a scoping review and stakeholder views.

Part III focuses on genetic diagnostics and implementation in ID care, which is a requisite for disorder-specific care. In **Chapter 8** the current situation of implementation of genetic diagnoses in multidisciplinary ID care is described and associated clinical and demographic factors to reveal possible health disparities are identified. Finally, in **Chapter 9** the findings of all research presented in this thesis are discussed, as well as next steps for the implementation of our findings and directions for future research. Recommendations are provided to foster clinical research for RGNDs and to enable personalized, disorder-specific care for these vulnerable affected individuals.

References

1. European Union. Regul. N°141/2000 Eur. Parliam. Counc. 16 December 1999 orphan Med. Prod. 2000.
2. Rode J. Rare diseases: understanding this public health priority. Paris: EURORDIS 2005.
3. Castrén E, Elgersma Y, Maffei L, Hagerman R. Treatment of Neurodevelopmental disorders in Adulthood. J. Neurosci. 2012;32(41).
4. Levy G, Barak B. Postnatal therapeutic approaches in genetic neurodevelopmental disorders [Internet]. 2021;3. Available from: https://doi.org/10.4103/1673-5374.293133.
5. Ellison JW, Rosenfeld JA, Shaffer LG. Genetic basis of intellectual disability. Annu. Rev. Med. 2013;64.
6. Leonard H, Wen X. The epidemiology of mental retardation: Challenges and opportunities in the new millennium. Ment. Retard. Dev. Disabil. Res. Rev. 2002;8(3).
7. Schalock RL, Luckasson R, Tassé MJ. An overview of intellectual disability: Definition, diagnosis, classification, and systems of supports (12th ed.). Am. J. Intellect. Dev. Disabil. 2021;126(6).
8. Maia N, Nabais Sá MJ, Melo-Pires M, et al. Intellectual disability genomics: current state, pitfalls and future challenges. BMC Genomics 2021;22(1).
9. Kvarnung M, Nordgren A. Intellectual disability & rare disorders: A diagnostic challenge. In: Advances in Experimental Medicine and Biology. 2017.
10. Cannizzo S, Lorenzoni V, Palla I, et al. Rare diseases under different levels of economic analysis: Current activities, challenges and perspectives. RMD Open 2018;4.
11. Kostanjsek N. Use of the International Classification of Functioning, Disability and Health (ICF) as a conceptual framework and common language for disability statistics and health information systems. In: BMC Public Health. 2011.
12. Paschos D. Intellectual Disability: Understanding its Development, Causes, Classifications, Evaluation, and Treatment. Child Adolesc. Ment. Health 2008;13(4).
13. Sappok T, Diefenbacher A, Winterholler M. The Medical Care of People With Intellectual Disability. Dtsch. Arztebl. Int. 2019.
14. Presson AP, Partyka G, Jensen KM, et al. Current estimate of down syndrome population prevalence in the United States. J. Pediatr. 2013;163(4).
15. von Scheibler ENMM, van Eeghen AM, de Koning TJ, et al. Parkinsonism in Genetic Neurodevelopmental Disorders: A Systematic Review. Mov. Disord. Clin. Pract. 2023;10(1).
16. Curatolo P, Moavero R, de Vries PJ. Neurological and neuropsychiatric aspects of tuberous sclerosis complex. Lancet Neurol. 2015.
17. Roth J, Roach ES, Bartels U, et al. Subependymal giant cell astrocytoma: Diagnosis, screening, and treatment. Recommendations from the international tuberous sclerosis complex consensus conference 2012. Pediatr. Neurol. 2013;49(6).
18. Huisman S, Mulder P, Kuijk J, et al. Self-injurious behavior. Neurosci. Biobehav. Rev. 2018;84:483–491.
19. Berg K, Arron K, Burbidge C, et al. Carer-Reported Contemporary Health Problems in People With Severe and Profound Intellectual Disability and Genetic Syndromes. J. Policy Pract. Intellect. Disabil. 2007;4(2).
20. De Villiers J, Porteous M. Genetic testing of adults with intellectual disability. Psychiatrist 2012;36(11)
21. Morant AV, Jagalski V, Vestergaard HT. Labeling of Disease-Modifying Therapies for Neurodegenerative Disorders. Front. Med. 2019.
22. Vernon HJ. Inborn errors of metabolism: Advances in diagnosis and therapy. JAMA Pediatr. 2015.
23. Jacquemont S, Berry-Kravis E, Hagerman R, et al. The challenges of clinical trials in fragile X syndrome. Psychopharmacology (Berl). 2014.

24. Jansen-Van Der Weide MC, Gaasterland CMW, Roes KCB, et al. Rare disease registries: Potential applications towards impact on development of new drug treatments. Orphanet J. Rare Dis. 2018.
25. Sun W, Zheng W, Simeonov A. Drug discovery and development for rare genetic disorders. Am. J. Med. Genet. Part A 2017;173(9).
26. Both P, ten Holt L, Mous S, et al. Tuberous sclerosis complex: Concerns and needs of patients and parents from the transitional period to adulthood. Epilepsy Behav. 2018.
27. Foster RH, Kozachek S, Stern M, Elsea SH. Caring for the caregivers: An investigation of factors related to well-being among parents caring for a child with smith-magenis syndrome. J. Genet. Couns. 2010.
28. Van Remmerden MC, Hoogland L, Mous SE, et al. Growing up with Fragile X Syndrome: Concerns and Care Needs of Young Adult Patients and Their Parents. J. Autism Dev. Disord. 2020.
29. De Kuijper G, Evenhuis H, Minderaa RB, Hoekstra PJ. Effects of controlled discontinuation of long-term used antipsychotics for behavioural symptoms in individuals with intellectual disability. J. Intellect. Disabil. Res. 2014.
30. Schoufour JD, Oppewal A, Van Der Maarl HJK, et al. Multimorbidity and polypharmacy are independently associated with mortality in older people with intellectual disabilities: A 5-year follow-up from the HA-ID study. Am. J. Intellect. Dev. Disabil. 2018;123(1).
31. Franz DN, Krueger DA. mTOR inhibitor therapy as a disease modifying therapy for tuberous sclerosis complex. Am. J. Med. Genet. Part C Semin. Med. Genet. 2018;178(3).
32. Hoytema van Konijnenburg EMM, Wortmann SB, Koelewijn MJ, et al. Treatable inherited metabolic disorders causing intellectual disability: 2021 review and digital app. Orphanet J. Rare Dis. 2021;16(1).
33. Kim J, Hu C, Moufawad El Achkar C, et al. Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. N. Engl. J. Med. 2019.
34. Nagree MS, Scalia S, McKillop WM, Medin JA. An update on gene therapy for lysosomal storage disorders. Expert Opin. Biol. Ther. 2019;19(7).
35. Overwater IE, Rietman AB, van Eeghen AM, de Wit MCY. Everolimus for the treatment of refractory seizures associated with tuberous sclerosis complex (TSC): Current perspectives. Ther. Clin. Risk Manag. 2019.
36. Curatolo P, Bjørnvold M, Dill PE, et al. The Role of mTOR Inhibitors in the Treatment of Patients with Tuberous Sclerosis Complex: Evidence-based and Expert Opinions. Drugs 2016;76(5).
37. Morrissey H, Pradhan T. Escitalopram in the treatment of anxiety in chromosome 18p deletion. J. Child Adolesc. Psychopharmacol. 2018;28(8).
38. Lott IT, Head E. Dementia in Down syndrome: unique insights for Alzheimer disease research. Nat. Rev. Neurol. 2019;15(3).
39. van der Poest Clement E, Jansen FE, Braun KPJ, Peters JM. Update on Drug Management of Refractory Epilepsy in Tuberous Sclerosis Complex. Pediatr. Drugs 2020.
40. Nikles J, Onghena P, Vlaeyen JWS, et al. Establishment of an International Collaborative Network for N-of-1 Trials and Single-Case Designs. Contemp. Clin. Trials Commun. 2021;23.
41. Abrahamyan L, Feldman BM, Tomlinson G, et al. Alternative designs for clinical trials in rare diseases. Am. J. Med. Genet. Part C Semin. Med. Genet. 2016;172(4).
42. Tate RL, Perdices M. Single-Case Experimental Designs for Clinical Research and Neurorehabilitation Settings. 2019.
43. OCEBM Levels of Evidence Working Group, Durieux N, Pasleau F, Howick J. The Oxford 2011 Levels of Evidence. Group 2011.
44. Guyatt G, Sackett D, Adachi J, et al. A clinician's guide for conducting randomized trials in individual patients. CMAJ 1988.
45. Shamseer L, Sampson M, Bukutu C, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015: Explanation and elaboration. BMJ 2015.
46. Lamont A, Lyons MD, Jaki T, et al. Identification of predicted individual treatment effects in randomized clinical trials. Stat. Methods Med. Res. 2018.

47. 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 Manage. 2015.
48. Bradbury J, Avila C, Grace S. Practice-Based Research in Complementary Medicine: Could N-of-1 Trials Become the New Gold Standard? Healthcare 2020.
49. Goetz LH, Schork NJ. Personalized medicine: motivation, challenges, and progress. Fertil. Steril. 2018;
50. Schork NJ. Personalized medicine: Time for one-person trials. Nature 2015.
51. Beckers LWME, Stal RA, Smeets RJEM, et al. Single-case Design Studies in Children with Cerebral Palsy: A Scoping Review. Dev. Neurorehabil. 2020.
52. Evans JJ, Gast DL, Perdices M, Manolov R. Single case experimental designs: Introduction to a special issue of Neuropsychological Rehabilitation. Neuropsychol. Rehabil. 2014.
53. Smith JD. Single-case experimental designs: A systematic review of published research and current standards. Psychol. Methods 2012.
54. Weinreich SS, Vrinten C, Kuijpers MR, et al. Aggregated N-of-1 trials for unlicensed medicines for small populations: an assessment of a trial with ephedrine for myasthenia gravis. Orphanet J. Rare Dis. 2017.
55. Hessl D, Sansone SM, Berry-Kravis E, et al. The NIH Toolbox Cognitive Battery for intellectual disabilities: Three preliminary studies and future directions. J. Neurodev. Disord. 2016;8(1).
56. Berry-Kravis EM, Hessl D, Rathmell B, et al. Effects of STX209 (Arbaclofen) on neurobehavioral function in children and adults with fragile X syndrome: A randomized, controlled, phase 2 trial. Sci. Transl. Med. 2012;4(152).
57. Jacquemont S, Curie A, Des Portes V, et al. Epigenetic modification of the FMR1 gene in fragile X syndrome is associated with differential response to the mGluR5 antagonist AFQ056. Sci. Transl. Med. 2011;3(64).
58. Budimirovic DB, Berry-Kravis E, Erickson CA, et al. Updated report on tools to measure outcomes of clinical trials in fragile X syndrome. J. Neurodev. Disord. 2017.
59. Shields RH, Kaat AJ, McKenzie FJ, et al. Validation of the NIH Toolbox Cognitive Battery in intellectual disability. Neurology 2020.
60. Dumont LJ, Aubuchon JP. Evaluation of proposed FDA criteria for the evaluation of radiolabeled red cell recovery trials. Transfusion 2008;48(6).
61. World Health Organization. Towards a Common Language for Functioning , Disability and Health ICF. Int. Classif. 2002.
62. Frew JW, Davidson M, Murrell DF. Disease-specific health related quality of life patient reported outcome measures in Genodermatoses: A systematic review and critical evaluation. Orphanet J. Rare Dis. 2017.
63. Thompson JJ, Kelly KL, Ritenbaugh C, et al. Developing a patient-centered outcome measure for complementary and alternative medicine therapies II: Refining content validity through cognitive interviews. BMC Complement. Altern. Med. 2011;11.
64. Valderas JM, Kotzeva A, Espallargues M, et al. The impact of measuring patient-reported outcomes in clinical practice: A systematic review of the literature. Qual. Life Res. 2008.
65. Mercieca-Bebber R, King MT, Calvert MJ, et al. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. Patient Relat. Outcome Meas. 2018.
66. Bijkerk CJ, De Wit NJ, Muris JWM, et al. Outcome measures in irritable bowel syndrome: Comparison of psychometric and methodological characteristics. Am. J. Gastroenterol. 2003;98(1).
67. Lee EH, Kwon O, Hahm KB, et al. Irritable bowel syndrome-specific health-related quality of life instrument: Development and psychometric evaluation. Health Qual. Life Outcomes 2016;14(1).
68. Van Eeghen AM, Bruining H, Wolf NI, et al. Personalized medicine for rare neurogenetic disorders: can we make it happen? Cold Spring Harb Mol Case Stud 2022;8(2):a006200.
69. Srikrishna S, Robinson D, Cardozo L, Gonzalez J. Is there a discrepancy between patient and physician quality of life assessment? Neurourol. Urodyn. 2009;28(3).

70. Zeltner NA, Baumgartner MR, Bondarenko A, et al. Development and psychometric evaluation of the metabQoL 1.0: A quality of life questionnaire for paediatric patients with intoxication-type inborn errors of metabolism. In: JIMD Reports. 2017.
71. Zöllner JP, Conradi N, Sauter M, et al. Quality of life and its predictors in adults with tuberous sclerosis complex (TSC): a multicentre cohort study from Germany. Neurol. Res. Pract. 2021;3(1).
72. Gaasterland CMW, Jansen-van der Weide MC, Vroom E, et al. The POWER-tool: Recommendations for involving patient representatives in choosing relevant outcome measures during rare disease clinical trial design. Health Policy (New. York). 2018.
73. 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.
74. Pekeles H, Accogli A, Boudrahem-Addour N, et al. Diagnostic Yield of Intellectual Disability Gene Panels. Pediatr. Neurol. 2019;92.
75. Wang J, Wang Y, Wang L, et al. The diagnostic yield of intellectual disability: Combined whole genome low-coverage sequencing and medical exome sequencing. BMC Med. Genomics 2020;13(1).
76. Blau N, Dionisi Vici C, Ferreira CR, et al. Correction to: Physician's Guide to the Diagnosis, Treatment, and Follow-Up of Inherited Metabolic Diseases. In: Physician's Guide to the Diagnosis, Treatment, and Follow-Up of Inherited Metabolic Diseases. 2022.
77. Kotulska K, Kwiatkowski DJ, Curatolo P, et al. Prevention of Epilepsy in Infants with Tuberous Sclerosis Complex in the EPISTOP Trial. Ann. Neurol. 2021;89(2).
78. Krueger DA, Wilfong AA, Holland-Bouley K, et al. Everolimus treatment of refractory epilepsy in tuberous sclerosis complex. Ann. Neurol. 2013;74(5).
79. van Karnebeek CDM, Jaggumantri S. Current Treatment and Management of Pyridoxine-Dependent Epilepsy. Curr. Treat. Options Neurol. 2015;17(2).
80. Savatt JM, Myers SM. Genetic Testing in Neurodevelopmental Disorders. Front. Pediatr. 2021;9
81. Baker K, Raymond FL, Bass N. Genetic investigation for adults with intellectual disability: Opportunities and challenges. Curr. Opin. Neurol. 2012;25(2).
82. Rojnueangnit K, Anthanont P, Khetkham T, et al. Genetic diagnosis for adult patients at a genetic clinic. Cold Spring Harb Mol Case Stud 2022;8(7):a006235.
83. Strydom A, Coppus A, Blesa R, et al. Alzheimer's disease in Down syndrome: An overlooked population for prevention trials. Alzheimer's Dement. Transl. Res. Clin. Interv. 2018;4.

PART I

N-of-1 studies in rare genetic neurodevelopmental disorders

Chapter 2

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, Kit C.B. Roes, Martina C. Cornel, Clara D.M. van Karnebeek, Frits A. Wijburg, Joost G. Daams, Erik Boot, Agnies M. van Eeghen

Neurology, 2021, *96*(11), 529-540.

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.

Introduction

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.

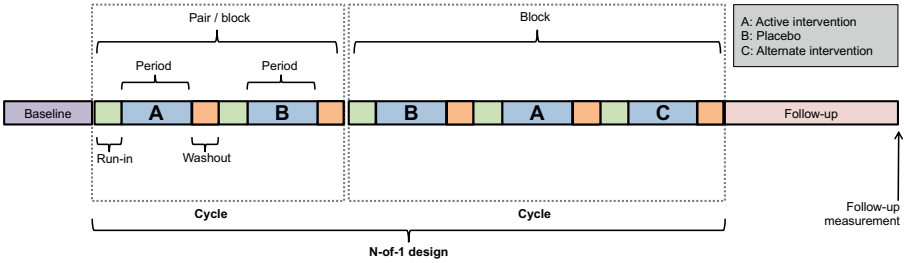


Figure 1. Schematic Presentation of Terminology Used for N-of-1 Trials.

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 two 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.

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 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 outcome measure(s) (presence and type), adverse events, power analysis,

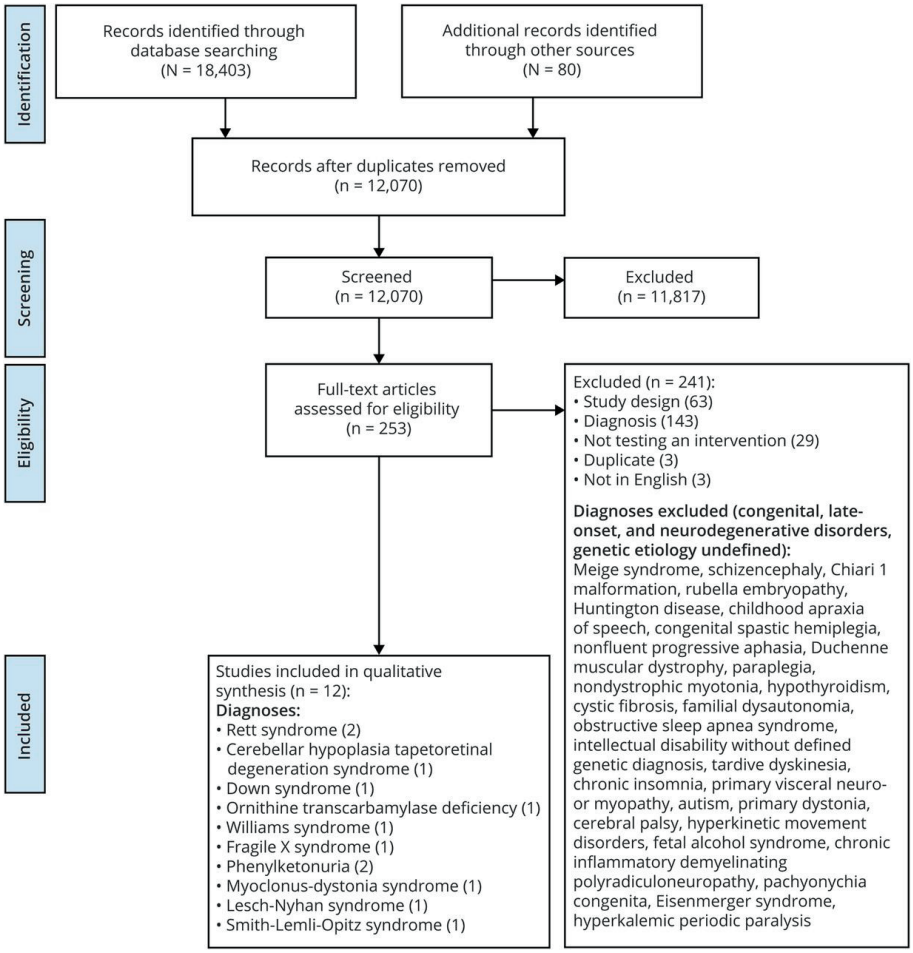


Figure 2. PRISMA Flowchart.
PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analysis.

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.

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 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 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.

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

Study	Diagnosis	Number of participants	Average age of participants (range); y	Intervention	Primary and secondary outcome measures[1]	Assessed by
Bawden et al. (55)	Williams syndrome	4	11 (9-13)	Methylphenidate	Child Behavior Checklist, Conners' Parent/ Teacher Questionnaire, Side Effects Questionnaire, cognitive psychometric measures	Caregiver
Byiers et al. (56)	Rett syndrome	3	30 (15-47)	Functional communication training	Communicative behaviour	Investigator
Camfield et al. (21)	Cerebellar hypoplasia tapetoretinal degeneration syndrome	6	7 (3-13)	Melatonin	Average number of hours asleep per 24 hours, and the number of awakenings and nights without arousals	Caregiver, 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, red blood cell folate levels	Caregiver, 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	L-arginine	Quality of life/mood assessment questionnaire, plasma glutamine and arginine levels	Patient, investigator
Khasnavis et al. (30)	Lesch-Nyhan disease	9	10 (6-22)	Ecopipam	Behavior Problems Inventory, Clinical Global Impression scale, adverse events	Caregiver, study staff
Luciano et al. (29)	Myoclonus-dystonia syndrome	2	29 (28-31)	Tetrabenazine	Global Dystonia rating scale, Fahn-Marsden rating scale, Unified Myoclonus rating scale	Investigator
Marholin et al. (25)	Phenylketonuria	6	36 (19-53)	Low phenylalanine diet and behavior modification	Social and motor behavior, 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

[1]Underlined when indicated as a primary outcome measure by the authors. y, year.

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

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

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.[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]

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]

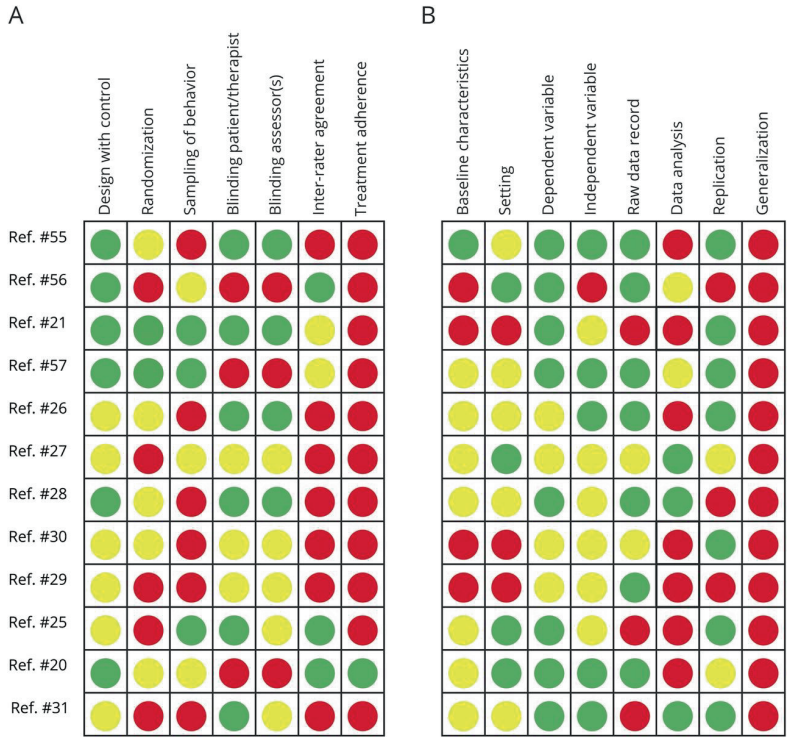


Figure 3. Schematic Representation of the Risk of Bias in N-of-1 Trials (RoBiNT) Scale.

(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).

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.

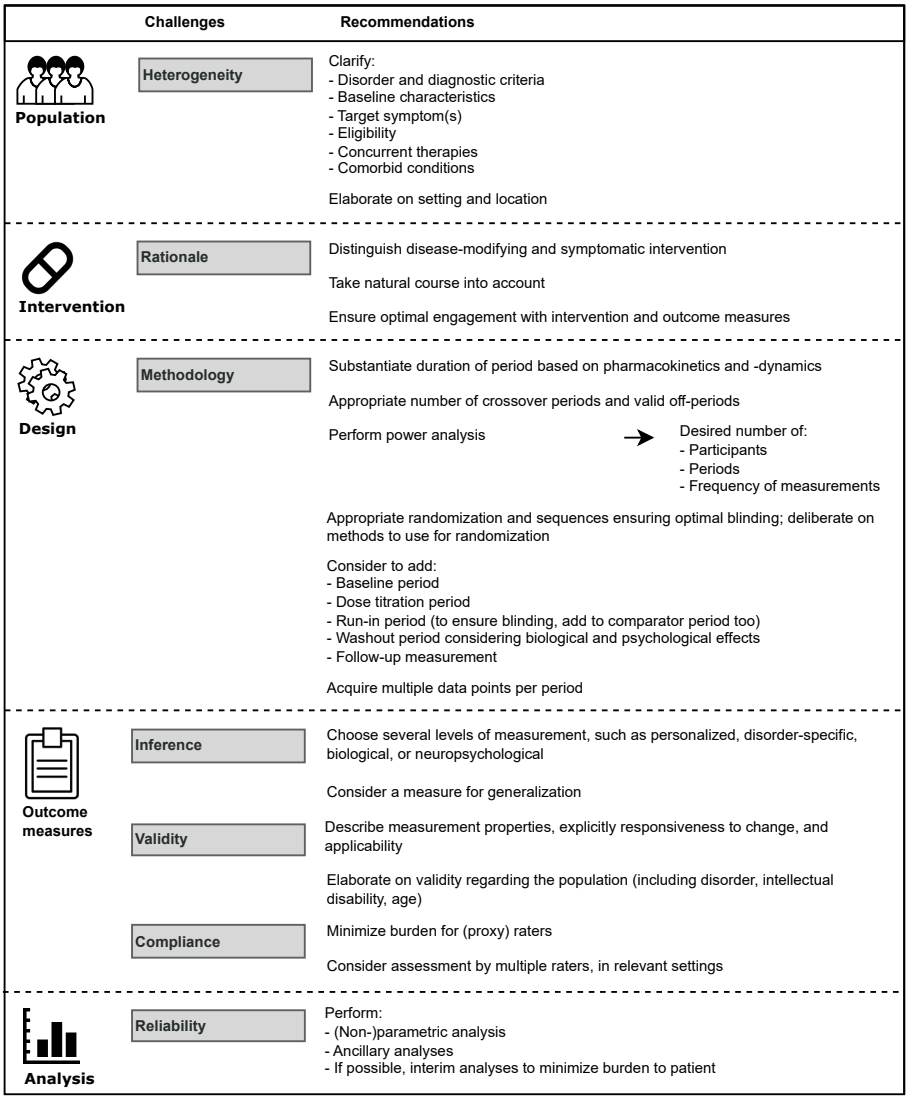


Figure 4. Challenges and Recommendations for Conducting and Reporting N-of-1 Studies in Rare Genetic Neurodevelopmental Disorders.

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 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 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 (Figure 4).[15,23] 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.

Conclusion

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.

Authors' contributions

AM was involved in the study concept and design, had a major role in the acquisition of data, analysis and interpretation of data, drafted the manuscript. MB was involved in the study concept and design, had a major role in the acquisition of data, revised the manuscript for content. PvdV and MC were involved in the study concept, had a major role in the acquisition of data, analysis or interpretation. KR was involved in the study concept, design, and analysis or interpretation of data. CvK and FW were involved in the study concept and design, and revised the manuscript for content. JD conducted the search strategy and contributed to the analysis and interpretation of data. EB was involved in the study concept and design, had a major role in the acquisition of data, analysis and interpretation, revised the manuscript for content, including medical writing. AvE was involved in the study concept and design, had a major role in the acquisition and interpretation of data, drafted and revised the manuscript for content, including medical writing.

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: https://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. https://www.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. Frontiers in medicine 2019;6:223.
4. Vernon HJ. Inborn errors of metabolism: Advances in diagnosis and therapy. JAMA Pediatrics 2015;169(8):778-782.
5. Guyatt G, Sackett D, Adachi J, et al. A clinician's guide for conducting randomized trials in individual patients. Canadian Medical Association Journal 1988;139(6):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: www.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. Statistical methods in medical research 2018;27(1):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. Journal of pain and symptom management 2015;50(3):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(4):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(1):15.
13. Schork NJ. Personalized medicine: Time for one-person trials. Nature 2015;520(7549):609-611.
14. Goetz LH, Schork NJ. Personalized medicine: motivation, challenges, and progress. Fertility and Sterility 2018;109(6):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. Neuropsychological rehabilitation 2013;23(5):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. Systematic reviews 2017;6(1):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. Systematic reviews 2016;4(1):1.
18. Van Karnebeek CDM, Stockler S. Treatable inborn errors of metabolism causing intellectual disability: A systematic literature review. Molecular Genetics and Metabolism 2012;105(3):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. Molecular Genetics and Metabolism 2014;111(4):428-438.

20. Simacek J, Dimian AF, McComas JJ. Communication Intervention for Young Children with Severe Neurodevelopmental Disabilities Via Telehealth. Journal of autism and developmental disorders 2017;47(3):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. Journal of Child Neurology 1996;11(4):341-343.
22. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Systematic reviews 2016;5(1):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 & dementia 2009;5(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. Pediatric Research 1978;12(3):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. American Journal of medical genetics 1988;30(1-2):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. Canadian Journal of Neurological Sciences 1980;7(2):127-131.
28. Hackett A, Gillard J, Wilcken B. n of 1 trial for an ornithine transcarbamylase deficiency carrier. Molecular genetics and metabolism 2008;94(2):157-161.
29. Luciano AY, Jinnah HA, Pfeiffer RF, Truong DD, Nance MA, LeDoux MS. Treatment of myoclonus-dystonia syndrome with tetrabenazine. Parkinsonism & related disorders 2014;20(12):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. Molecular genetics and metabolism 2016;118(3):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. American journal of medical genetics Part A 2010;152(1):91-95.
32. Gupta S, Faughnan ME, Tomlinson GA, Bayoumi AM. A framework for applying unfamiliar trial designs in studies of rare diseases. Journal of clinical epidemiology. 2011;64(10):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. Journal of the American Medical Association 2000;284(10):1290-1295.
34. Curatolo P, Moavero R. mTOR Inhibitors in Tuberous Sclerosis Complex. Current neuropharmacology 2012;10(4):404-415.
35. Shapiro EG, Escolar ML, Delaney KA, Mitchell JJ. Assessments of neurocognitive and behavioral function in the mucopolysaccharidoses. Molecular Genetics and Metabolism 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. Journal of Intellectual Disability Research 2004;48(2):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. Journal of clinical epidemiology 2011;64(5):471-480.
38. Ballard C, Mobley W, Hardy J, Williams G, Corbett A. Dementia in Down's syndrome. The Lancet Neurology 2016;15(6):622-636.
39. Van Eeghen AM, Black ME, Pulsifer MB, Kwiatkowski DJ, Thiele EA. Genotype and cognitive phenotype of patients with tuberous sclerosis complex. European Journal of Human Genetics 2012;20(5):510-515.

40. Araujo A, Julious S, Senn S. Understanding variation in sets of N-of-1 trials. PLoS One Epub 2016 Dec 1;11(12):e0167167.
41. Senn S. Mastering variation: Variance components and personalised medicine. Statistics in medicine 2016;35(7):966-977.
42. Schmid, C. H., & 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 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(12):1287-1294.
44. Slade A, Isa F, Kyte D, et al. Patient reported outcome measures in rare diseases: A narrative review. Orphanet journal of rare diseases 2018;13(1):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 journal of rare diseases 2015;10(1):59.
46. Ettema TP, Dröes RM, Lange J De, Mellenbergh GJ, Ribbe MW. A review of quality of life instruments used in dementia. Quality of life research 2005;14(3):675-686.
47. Harris SL, Handleman JS, Belchic J, Glasberg B. The vineland adaptive behavior scales for young children with autism. Special Services in the Schools 1995;10(1):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. Research in developmental disabilities 2013;34(5):1804-1814.
49. Shields RH, Kaat AJ, McKenzie FJ, et al. Validation of the NIH Toolbox Cognitive Battery in intellectual disability. Neurology 2020;94(12):1229-1240.
50. O'Mahony B, Skinner MW, Noone D, Page D, O'Hara J. Assessments of outcome in haemophilia - a patient perspective. Haemophilia: the Official Journal of the World Federation of Hemophilia 2016;22(3):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 Medical Research Methodology 2019;19(1):227.
52. Csikszentmihalyi M, Larson R. Validity and reliability of the experience-sampling method. In: Flow and the foundations of positive psychology 2014;35-54.
53. Senn S. Sample size considerations for n-of-1 trials. Statistical Methods in Medical Research 2019;28(2):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.
55. Bawden HN, MacDonald GW, Shea S. Treatment of children with Williams syndrome with methylphenidate. Journal of Child Neurology 1997;12(4):248-252.
56. Byiers BJ, Dimian A, Symons FJ. Functional communication training in rett syndrome: A preliminary study. Americal Journal on Intellectual and Developmental Disabilities 2014;119(4):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. Journal of Applied Research in Intellectual Disabilities 2016;29(1):1-10.

Chapter 3

Methylphenidate for attention-deficit/hyperactivity disorder in patients with Smith-Magenis syndrome: protocol for a series of N-of-1 trials

Annelieke R. Müller, Janneke R. Zinkstok, Nanda N.J. Rommelse, Peter M. van de Ven, Kit C.B. Roes, Frits A. Wijburg, Ester de Rooij-Askes, Cathelijne Linders, Erik Boot, Agnies M. van Eeghen

Orphanet Journal of Rare Diseases, 2021, *16*(1), 1-9.

Abstract

Background: Smith–Magenis syndrome (SMS) is a rare genetic neurodevelopmental disorder characterized by intellectual disability and severe behavioural and sleep disturbances. Often, patients with SMS are diagnosed with attention-deficit/hyperactivity disorder (ADHD). However, the effectiveness of methylphenidate (MPH), the first-line pharmacological treatment for ADHD, in patients with SMS is unclear. Our objective is to examine the effectiveness of MPH for ADHD symptoms in individuals with SMS, proposing an alternative trial design as traditional randomized controlled trials are complex in these rare and heterogeneous patient populations.

Methods and analysis: We will initiate an N-of-1 series of double-blind randomized and placebo-controlled multiple crossover trials in six patients aged ≥ 6 years with a genetically confirmed SMS diagnosis and a multidisciplinary established ADHD diagnosis, according to a power analysis based on a summary measures analysis of the treatment effect. Each N-of-1 trial consists of a baseline period, dose titration phase, three cycles each including randomized intervention, placebo and washout periods, and follow-up. The intervention includes twice daily MPH (doses based on age and body weight). The primary outcome measure will be the subscale hyperactivity/inattention of the Strengths and Difficulties Questionnaire (SDQ), rated daily. Secondary outcome measures are the shortened version of the Emotion Dysregulation Inventory (EDI) reactivity index, Goal Attainment Scaling (GAS), and the personal questionnaire (PQ). Statistical analysis will include a mixed model analysis. All subjects will receive an assessment of their individual treatment effect and data will be aggregated to investigate the effectiveness of MPH for ADHD in SMS at a population level.

Conclusions: This study will provide information on the effectiveness of MPH for ADHD in SMS, incorporating personalized outcome measures. This protocol presents the first properly powered N-of-1 study in a rare genetic neurodevelopmental disorder, providing a much-needed bridge between science and practice to optimize evidence-based and personalized care.

Trial registration

This study is registered in the Netherlands Trial Register (NTR9125).

Highlights of the study protocol

- Innovative trial design combining collection of scientific data with personalized care, providing a much-needed bridge between practice and science.
- Evidence-based treatment of ADHD symptoms in Smith–Magenis syndrome.
- The first adequately powered series of randomized, double-blind, placebo-controlled N-of-1 trials for a rare genetic neurodevelopmental disorder.
- Exploring patient-centered outcome measures addressing relevant goals of the patient.

Introduction

Smith–Magenis syndrome (SMS) is a rare genetic neurodevelopmental disorder with an estimated prevalence of 1:15.000–25.000 births.[1] SMS is caused by a deletion on chromosome 17 (17p11.2) or a pathogenic mutation in the *RAI1* gene located within this region. Most of the SMS manifestations are due to haploinsufficiency of *RAI1* and thought to be modified by other genes in the 17p11.2 region.[2,3,4] Manifestations are variable and include intellectual disability (ID), severe sleep disturbances and psychiatric comorbidity such as autism spectrum disorders (ASD), attention-deficit-hyperactivity disorder (ADHD).[5,6,7] Typical behavioural manifestations include problems with emotion dysregulation, self-injurious behaviour and aggressive or stereotypical behaviour, posing a great burden on patients and caregivers.[8]

Treatment of the behavioural manifestations in SMS is complex due to the genetic heterogeneity, clinical variability and severity of symptoms.[4,9] Traditionally, treatment is focused on appropriate management of sleeping pattern, concomitant somatic comorbidities, psycho-education and professional guidance for parents and caregivers aimed at symptom reduction and optimizing quality of life of both the patient and their family.[10,11,12,13] Often, this does not suffice, resulting in the prescription of psychotropic drugs in the vast majority of children and adults with SMS, including stimulants, antidepressants, antipsychotics, mood stabilizers, alfa2 agonists, sleep aids, and benzodiazepines.[14]

For idiopathic ADHD, methylphenidate (MPH) is well-established as first-line treatment with high efficacy and tolerability compared to other psychotropic drugs.[15,16,17] However, for ADHD in genetic neurodevelopmental disorders such as SMS more information is necessary as there is increasing evidence for differential treatment response and tolerability.[14,18,19] Also, polypharmacy is a clinical pitfall in patients with complex psychiatric disorders and ID, leading to iatrogenic comorbidity.[20] Therefore, disorder-specific studies are needed to provide information about the effectiveness of MPH for ADHD. Considering the heterogeneity of the patient population and need for relevancy of interventions, personalized outcome measures are needed to

enable measurement of clinically important changes. Such a personalized methodological approach has the potential of maximizing treatment adherence that is both patient-centered and evidence-based.[21,22,23]

Rationale for N-of-1 design

Trials in rare genetic neurodevelopmental disorders such as SMS pose specific challenges due to comorbidities and rarity of conditions.[24,25] Single-case experimental designs (SCEDs) provide an alternative to traditional parallel group randomized controlled trials (RCTs). Of SCEDs, the N-of-1 methodology provides the most rigorous evidence for treatment decisions at an individual level as replication is key for confirmation of causality. N-of-1 studies are randomized, controlled, multiple cross-over trials within individual patients[26,27] and enhance precision when treatment effects are heterogeneous between individuals.[28,29] Aggregating the results of several N-of-1 trials potentially yields treatment effect estimates that may be generalized at population level and may be as robust as traditional RCTs.[30] In particular, patients with rare disorders require individualized treatment interventions and outcomes due to their heterogeneity and vulnerability, which is facilitated by N-of-1 designs and consistent with the movement towards personalized care, providing a much needed bridge between practice and science.[21]

Objectives

The main objective is to study the effectiveness of MPH for ADHD symptoms in individuals with SMS. Secondary objectives include assessment of the effect of MPH on emotion dysregulation, personalized goals that are specific and important to the patient, and side effects. To do this, we will perform a series of N-of-1 trials as these provide an excellent approach to study effectiveness of MPH on ADHD in SMS, given: (1) the chronic and relatively stable clinical course of ADHD, and (2) the rapid onset and termination of action of MPH.[31]

Methods

Study design

We used the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) extension for N-of-1 trials (SPENT) checklist that is aligned with the CONSORT (consolidated reporting items for trials) extension for N-of-1 trials (CENT) for developing this N-of-1 protocol.[29]

The study will consist of a series of N-of-1 trials followed by an optional open-label extension phase. Each trial is randomized, placebo-controlled, and double-blinded with multiple crossovers within a single patient. The trial consists of a baseline period, dose titration phase, and three cycles each consisting of one period of MPH treatment and one period of placebo treatment, both followed by a one-week washout period (Figure 1). Despite the fact that a one-day washout would suffice biologically, we chose one-week washouts to account for prolonged psychological effects that may occur. The order of the treatment periods will be randomized. Thus, each N-of-1 trial will last 14 weeks with an additional follow-up measurement three months after completion of the N-of-1 trial.

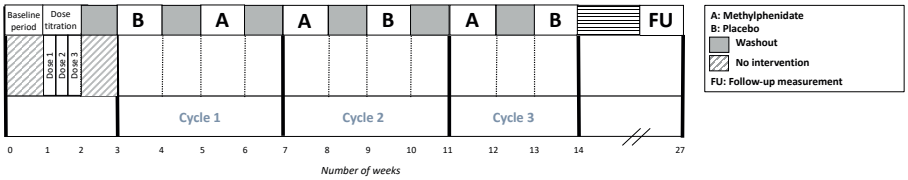


Figure 1. Study design.

Protocol development and patient engagement

Collaboration with the Dutch SMS patient advocacy organization, caregivers of patients and clinical experts played a large role in defining knowledge and care gaps, prioritizing the treatment study, development of the current protocol and selecting outcome measures. We addressed specific difficulties for conducting this study, including concerns related to caregiver burden and patient burden of participation, and issues for recruitment and retention.

Outcome measures

The primary outcome is the change on the hyperactivity/inattention subscale of the Strengths and Difficulties Questionnaire (SDQ) during active interventional periods. Secondary outcome measures are the shortened version of the Emotion Dysregulation Inventory (EDI) reactivity index,[32] Goal Attainment Scaling (GAS)[33] and the personal questionnaire (PQ).[34] Also, (the number of) side effects determined by the side effects checklist of MPH will be recorded.

Rationale for outcome measures

The SDQ subscale and the shortened version of the EDI have both been psychometrically considered as valid tools to measure behavior of people with ID and applicable to both children and adults.[32,35,36] Specifically, the SDQ was found to be a valid outcome measure for children with ADHD symptoms and showed preliminary results of validation for children with ID.[37,38] EDI was created using methods developed by the Patient-Reported Outcomes Measurement Information System (PROMIS) and validated as an efficient and sensitive method to measure emotion dysregulation in youth with ASD of any level of cognitive or verbal ability.[32,36] The EDI will serve as a generalization measure that is defined as an outcome closely or more distally related to the target behavior, and is used to evaluate transfer effects of the intervention to a broader domain of functioning.[39] For instance, it could be the same behavior but in another setting, such as inattention at school and at home, or interventional effects on a completely different behavior, such as improved emotion regulation when the target behavior is impulsivity. In addition to the target behaviors hyperactivity and inattention in our study, measured by the SDQ, MPH might affect emotion dysregulation as well, which could be measured by the EDI. GAS is an individualized outcome measure involving goal selection and goal scaling that is standardized in order to calculate the extent to which a patient's goals are met. Patients and/or their caregivers are allowed to choose their own specific goals in coordination with their treating physician/therapist. This makes GAS a measurement instrument that is very sensitive to change, particularly in small heterogeneous groups.

As the population with ID often presents with atypical side effects, a standardized checklist of side effects of MPH[40] together with an open interview to capture possible atypical side effects will be used to determine (the number of) side effects including sleeping problems.

Study population

The study population consists of children or adults from the Netherlands with SMS and an ADHD diagnosis established by a multidisciplinary team. Inclusion criteria are a minimum of six years old, a genetically confirmed diagnosis of SMS, and the availability of a caregiver for proxy-reports. Baseline characteristics will be recorded in detail, including age, gender, genetic test results, comorbidity, and medication. Exclusion criteria include presence of a contra-indication for MPH, planned general anesthesia, pregnancy, breastfeeding, current treatment with biologically interfering drugs, substance or alcohol abuse, and incapacity to swallow tablets. The latter may however bias the sample toward a higher functioning segment of SMS. We aim to conduct a patient-centered trial, allowing for a natural setting and flexibility, including the continuation of concurrent therapies such as (for example) sleep medication. Use of concurrent therapies will be recorded.

Sample size

The sample size calculation was based on a summary measures analysis of the treatment effect as measured with the primary outcome SDQ.[41] The difference between the mean SDQ hyperactivity/inattention ratings in MPH periods and placebo periods was used as a summary measure for the treatment effect in an individual subject. The estimated standard deviation (SD) of 2.3 points for single ratings was used based on a reported standard error for the parent-rated SDQ subscale.[42] Using a test–retest intraclass correlation coefficient (ICC) of 0.84,[43] we decomposed a SD into a within-subject SD of 0.92 and a between-subject SD of 2.11. Assuming an SD of 1 point for the treatment effect, 95% of the subject-specific treatment effects roughly falls within a range of 4 points. Based on the estimate assuming three cycles with seven daily SDQ ratings within each period, a total of 6 subjects will yield 80% power to detect a mean difference of 1.5 points

between intervention and placebo periods when assuming a two-sided significance level of 5%.

Recruitment

Study subjects will be recruited through the two national Dutch SMS multidisciplinary outpatient clinics of 's Heeren Loo, and the Dutch SMS patient advocacy organization.

Trial procedure and study setting

Prior to the start of the trials, the participant and substitute decision maker(s) will have a clinical visit to discuss the procedure in detail and sign the informed consent. Personalized goals with regard to GAS and the PQ and target symptoms will be identified together by the parents and/or primary caregivers, the treating physician, psychologist and/or behavioural therapist, and investigator. During the clinical visit, it will be emphasized that assessors should rate the global effect over the day and should be aware of the possible rebound effect of MPH. The study will be carried out at participants' home setting and schools or daytime centres if applicable.

The trial will start with a baseline period of seven days without any intervention. A dose titration phase of six days is followed by a washout period of eight days. The individual N-of-1 trial will consist of three cycles each containing four seven-day periods: one active treatment (A), one placebo treatment (B), and two 'washout' periods following A and B. The order in which patients receive active and placebo treatment is randomized within each cycle. The medication will be administered at home and/or at school or daytime activities by parents or primary caregivers. During the baseline period and three cycles, the SDQ and EDI will be filled out daily at the end of the day using app-based questionnaires by primary caregivers (Figure 2). Filling out the questionnaires will take about 1 min a day. At the end of each seven-day period, the investigator will interview patients and/or primary caregivers by phone to evaluate goals,[33] to assess possible side effects, to note the general moments that the interventional effects seem to wear off, and to note the perceived treatment received (MPH or placebo). The time expected to complete this interview is 15 min. Each period will include a weekend such that parents can provide assessments of complete

days. At the end of the trial period, the participant will have a second and final clinical visit to evaluate the symptoms and study. In consultation with the treating physician, patients may continue with MPH treatment, whether or not at a different dosage. Three months after terminating the N-of-1 trial, another contact moment will take place for a follow-up measurement in which the questionnaires will be filled out and the goals and items of GAS and PQ will be discussed again. To reduce burden as much as possible, assessments solely occur by phone calls apart from the two study visits. The total duration of the trial will be 14 weeks with the additional follow-up measurement after three months.

		Enrollment*	Allocation	Baseline	Dose titration	Washout	Cycle 1	Cycle 2	Cycle 3*	Follow-up
Time point (weeks)				**0**	**1**	**2**	**3-6**	**7-10**	**11-14**	**27**
Enrolment	Eligibility screen	X								
	Informed consent	X								
	Allocation		X							
Interventions	Methylphenidate				X		X	X	X	X
	Placebo						X	X	X	X
Assessments	SDQ (daily)			X			X	X	X	X
	EDI (daily)			X			X	X	X	X
	GAS	X		X			X	X	X	X
	PQ	X		X			X	X	X	X
	Side effects				X		X	X	X	X

Figure 2. Time schedule of enrolment, interventions, and assessments.

Underlined crosses (X) indicate assessments via phone calls. Asterisks (*) indicate the moment with a clinical visit. EDI Emotion Dysregulation Inventory, GAS Goal Attainment Scaling, PQ personal questionnaire, SDQ Strengths and Difficulties Questionnaire.

Blinding, treatment allocation, randomization

Participants, parents, caregivers, supervisors of daily activities, clinicians and researchers will all be blinded during the N-of-1 trial. The random allocation sequence will be generated and implemented by the hospital pharmacist for block randomization in a 1:1 ratio and sequentially numbered packages. Participants and the treating physician will be deblinded after

completing the three cycles or in case of serious adverse events (SAEs). Investigators involved in data analysis will remain blinded until the end of the follow-up period.

Multi-site training plan

A pre-study training meeting will be planned to train clinical investigators and clinical evaluators on study procedures and GAS with a secondary goal to promote reliability of GAS. All clinical and research staff that is involved in either identification or assessment of goals by GAS will be trained by a GAS expert to promote data quality.

Interventions and dosing schedule

One dose titration kit and a trial kit including MPH (regular tablet) and placebo will be developed and distributed by the Amsterdam UMC hospital pharmacist.

Dose titration phase

The MPH dosage will be titrated to achieve the maximum dosage with minimal side effects determined by the psychiatrist or ID physician. Titration dosage will be blinded to the participants and caregivers and comprise two days each of three escalating doses in steps of 2.5 mg of MPH with a total of six days followed by a washout period of at least one week. The individually determined starting dose for the dose titration phase will be based on age and body weight. During the dose titration phase, participants will daily fill out the checklist of side effects of MPH.[40] MPH effectiveness will explicitly not be examined during the titration phase to prevent high dropout rates when participants might get prematurely convinced about the effectiveness.

Trial

During the N-of-1 trial, MPH dosage as determined by titration phase or placebo will be administered by caregivers twice daily during breakfast and during lunch (around 7.30 am and 12.30 pm). During washout periods, the placebo will be administered.

Follow-up

After the final cycle and unblinding, the participant's substitute decision maker(s) and clinician will decide on further continuation of MPH treatment before the follow-up measurement. Although a dose titration phase precedes the trial to have a fixed dosage during the N-of-1 trial, participants can switch from dosage or discontinue with MPH in consultation with the treating physician in the follow-up period.

Safety evaluation

Subjects can leave the study at any time for any reason. The investigator may decide to withdraw a subject from the study for urgent medical reasons. Reasons may include occurrence of treatment-related SAEs or suspected unexpected serious adverse reaction (SUSAR), deterioration of symptoms that require a treatment other than the medication of the trial, and a sudden and acute medical condition related or unrelated to SMS that may interfere with the study. Any sign that indicates resistance among children and mentally incompetent participants, which is defined and discussed with parents and caregivers in advance, will lead to discontinuation of the trial. Completed cycles before withdrawal of a participant will still be analysed. In case of drop-out, a new participant that meets the inclusion criteria will be recruited with a newly randomized sequence. The sponsor will suspend the study if there is sufficient ground that continuation of the study will jeopardise subject health or safety.

Monitoring will be conducted by independent qualified monitors from the Clinical Monitoring Center (CMC). All adverse events (AEs) will be monitored and followed until they have abated, or until a stable situation has been reached. Depending on the event, follow-up may require additional tests or medical procedures as indicated, and/or referral to the general practitioner or a medical specialist.

Data collection and management

All data will be collected and handled in accordance with the EU General Data Protection Regulation, the Dutch Act on Implementation of the General Data Protection Regulation and Amsterdam UMC standard operating procedures. The Case Report Forms (CRFs) and trial specific documents

held by the researcher will be stored securely with access restricted and limited to nominated research staff recorded on the delegation log. A data sharing agreement between Amsterdam UMC and 's Heeren Loo will manage additional access for investigators.

The CRFs will be set up in Castor Electronic Data Capture (EDC) in which weekly assessments will be entered. Questionnaires can be filled out digitally using the m-Path app on smartphones,[44] on computers (Castor EDC) or by using paper forms. Data from the app will be collected at the end of each trial and will be loaded into Castor EDC. In advance, participants will be recommended to download the m-Path app to easily and confidentially answer the daily questionnaires, although the use of different ways is allowed to enlarge feasibility for raters. For the sake of participant retention, automatic reminders will be sent to raters when questionnaires have not yet been filled in. Participant burden will be limited as much as possible by having contact moments by video-conference or phone instead of a visit. The investigator can also decide to withdraw a subject for urgent medical reasons. A participant who withdraws consent for an assessment of one outcome may be willing to continue with assessments for other outcomes.

A subject identification code list will be used with unique participant identifiers not deducible to patients. Only two investigators will have access to the key. In addition, two methodologists and biostatisticians will have access to the source data for methodological and statistical purposes. Data will be stored for 15 years according to the Amsterdam UMC regulations.

Statistical methods

An individual treatment effect for each participant will be determined based on summary statistics. A mixed model analysis will be applied for analysing the effectiveness of the intervention at the population level combining data from the individual N-of-1 trials.

The mean treatment effect on the primary outcome will be estimated and tested for significance using a linear mixed model with a fixed effect for treatment (MPH or placebo) and random effects for patient, cycle within patient, and treatment (within patient). The mixed model will account for

between-subjects heterogeneity in treatment effect through inclusion of the random treatment effect. Small amounts of missing data will not pose problems for the mixed model analysis because of the many data points per period, assuming data is missing at random. If issues such as singularity arise due to complexity of the models, an analysis based on a summary measure will be performed. A similar method will be used for estimating treatment effects on secondary study parameters. A two-sided significance level of 5% will be used. Analyses will be performed in R, using the lmer package.

Discussion

To date, research on the efficacy of treatment strategies for behavioural aspects of SMS has been limited. In this N-of-1 series of randomized, placebo-controlled, double-blind multiple crossover trials in patients with SMS and ADHD, the effectiveness of MPH for ADHD symptoms will be examined, including personalized goals as additional outcomes.

N-of-1 studies provide a powerful alternative to larger RCTs, but are still only sporadically reported in rare genetic neurodevelopmental disorders.[45] Debate is still ongoing to what extent an N-of-1 study represents medical research or is part of evidence-based clinical care.[46,47,48] For instance, for some practitioners starting MPH treatment, blinded crossover periods, the use of placebo and filling out questionnaires is already part of standard care. To provide evidence-based treatment decisions and to prevent polypharmacy, N-of-1 studies might be considered as a much-needed part of clinical care especially in complex patient populations such as individuals with SMs.

Combining personalized and relevant treatment targets while pursuing optimal generalizability is challenging in heterogeneous patient populations such as SMs. Because SMs is accompanied by various and often variable levels of ID and comorbidities, clear diagnostic and eligibility criteria are necessary and baseline characteristics, concurrent therapies, comorbid conditions and target symptoms will be clearly defined to optimize

interpretation and generalizability. Also, we will elaborate on setting and location as assessments will be in the participant's natural environment.

Regarding this symptomatic pharmacological intervention, we chose to add a baseline period. This period allows us to observe the behavior in a non-clinical trial setting and to take the natural course of ADHD symptoms into account. Moreover, to ensure optimal efficacy, tolerability and hence compliance, the highest dosage without side effects will be chosen based on the dose titration phase.
As for the design, the number of participants and crossover periods to detect a clinically relevant treatment effect was selected based on a power analysis, providing the first properly powered N-of-1 study in a rare genetic neurodevelopmental disorder.[41] These are needed when intending to provide estimates of the treatment effect at a population level. Duration of periods was based on the pharmacokinetics and -dynamics of MPH. Although no washout period would suffice pharmacologically, one-week washouts were chosen to account for prolonged psychological effects and for planning purposes.

To pursue optimal generalizability to the entire SMS population, it is of great importance that outcome measures are validated for the patient population and sensitive to change. Multiple data points per period will be acquired to enable estimation of between and within-period variances. To increase the study's validity, each interventional period includes at least five measurements of the target symptoms, by using the subscale of the SDQ.[26,49] Several other domains of measurement were chosen, such as sleep quality and personalized measurements. GAS also allows for capturing goals in reduction of caregiver stress, as reduction in symptoms may have benefit for family as well. The EDI will also serve as a generalization measure to evaluate transfer effects of the intervention to a broader domain of functioning. Generalization measures are dependent variables that are taken in addition to the target behavior that are used to evaluate whether an intervention generalizes to other behaviors or settings.[39] A shortened version and a subscale of two outcome measures were selected to minimize assessor's burden.

Personalized outcome measures such as GAS and the PQ were chosen to appraise subjective experiences in daily life, enabling quantitative expression of meaningful subjective patient experiences while translating these into evidence.[50] Trials tailored to participants by using personalized outcomes may improve treatment adherence as well. Although GAS has not yet been validated and performed in N-of-1 designs nor as an outcome measure in rare genetic disorders with ID, it may be a valuable tool in a complex and heterogeneous population such as SMS. This study will introduce GAS in the N-of-1 design and might be a step towards validation of this personalized outcome measure in rare disorders.

Regarding the analysis, a mixed model analysis was selected to analyze the effectiveness of the intervention at the population level, accounting for between-subjects heterogeneity. Ancillary analyses will be performed to evaluate period effects and intrasubject correlation.

Limited burden is expected and maximal relevance and treatment adherence is ensured, as an N-of-1 study provides the unique opportunity to tailor interventions and outcomes to individual patients. To optimize compliance, daily questionnaires will be filled out using a user-friendly app and contact moments will mainly take place via digital or telephone calls. Caregivers may experience some burden because of longer withholding of active medication due to one-week washouts to account for eventual psychological effects; this was also the main aberrance from clinical care, necessitating institutional review board (IRB)-approval. On the other hand, every participant is exposed to the active treatment condition and the effect of the individual treatment will be assessed in the best available way, minimizing placebo effects, observer effects, and confirmation biases. After the N-of-1 trial, participants and their representative(s) will be fully informed on the effectiveness of the intervention, allowing shared decision making on future treatment. Participants might thus be particularly motivated to participate in an N-of-1 study due to the existing paucity of evidence and the fact that all subjects will receive an evidence-based assessment of their individual treatment effect.

Conclusion

This N-of-1 study will allow the delivery of personalized care while acquiring evidence of MPH for ADHD in the SMS population. We expect that use of the N-of-1 methodology and patient-centered outcome measures will assist in realizing the urgently needed evidence-based interventions in patients with rare genetic neurodevelopmental disorders. This protocol will be applicable for other genetic syndromes, and more N-of-1 series will allow cross-disorder comparisons and investigation of generalizability to the whole population with these disorders and/or ID. This study protocol can be used as a model to empower other clinician-researchers to investigate much-needed symptomatic pharmacological as well as disease-modifying interventions in rare disorders using a collaborative and multi-disciplinary approach.

Acknowledgements

The authors would like to thank the Dutch SMS patient advocacy organization (Stichting Smith Magenis Syndroom Nederland) and Charlotte Gaasterland and Hanneke van der Lee for their feedback on the study protocol.

Funding

The trial is financially sponsored by the Amsterdam UMC and health care institution 's Heeren Loo.

Competing interests

The authors declare that they have no competing interest.

Authors' contributions

AM, EdR and AvE proposed the study and initiated the design. AM and AvE wrote the study protocol with input from NR, EB, JZ, and FW. KR provided methodological expertise and PvdV designed the statistical analysis and will conduct the analyses. AM and AvE drafted the manuscript. All authors read and approved the final manuscript.

References

1. Elsea SH, Williams SR. Smith-Magenis syndrome: Haploinsufficiency of RAI1 results in altered gene regulation in neurological and metabolic pathways. Expert Rev Mol Med. 2011.
2. Boot E, Linders CC, Tromp SH, van den Boogaard MJ, van Eeghen AM. Possible underreporting of pathogenic variants in RAI1 causing Smith–Magenis syndrome. American Journal of Medical Genetics, Part A. 2021.
3. Edelman EA, Girirajan S, Finucane B, Patel PI, Lupski JR, Smith ACM, et al. Gender, genotype, and phenotype differences in Smith-Magenis syndrome: A meta-analysis of 105 cases. Clin Genet. 2007;71(6).
4. Elsea SH, Girirajan SS. Smith-Magenis syndrome. Eur J Hum Genet. 2008.
5. Laje G, Morse R, Richter W, Ball J, Pao M, Smith ACM. Autism spectrum features in Smith-Magenis syndrome. Am J Med Genet Part C Semin Med Genet. 2010.
6. Poisson A, Nicolas A, Cochat P, Sanlaville D, Rigard C, De Leersnyder H, et al. Behavioral disturbance and treatment strategies in Smith-Magenis syndrome. Orphanet J Rare Dis. 2015.
7. Greenberg F, Guzzetta V, De Oca-Luna RM, Magenis RE, Smith ACM, Richter SF, et al. Molecular analysis of the Smith-Magenis syndrome: A possible contiguous- gene syndrome associated with del(17)(p11.2). Am J Hum Genet. 1991.
8. Sloneem J, Oliver C, Udwin O, Woodcock KA. Prevalence, phenomenology, aetiology and predictors of challenging behaviour in Smith-Magenis syndrome. J Intellect Disabil Res. 2011.
9. Shayota BJ, Elsea SH. Behavior and sleep disturbance in Smith-Magenis syndrome. Current Opinion in Psychiatry. 2019.
10. Shelley BP, Robertson MM. The neuropsychiatry and multisystem features of the Smith-Magenis Syndrome: A review. Journal of Neuropsychiatry and Clinical Neurosciences. 2005.
11. Fidler DJ, Hodapp RM, Dykens EM. Stress in families of young children with down syndrome, williams syndrome, and smith-magenis syndrome. Early Educ Dev. 2000.
12. Hodapp RM, Fidler DJ, Smith ACM. Stress and coping in families of children with Smith-Magenis syndrome. J Intellect Disabil Res. 1998.
13. Nag HE, Hoxmark LB, Nærland T. Parental experiences with behavioural problems in Smith–Magenis syndrome: The need for syndrome-specific competence. J Intellect Disabil. 2019.
14. Laje G, Bernert R, Morse R, Pao M, Smith ACM. Pharmacological treatment of disruptive behavior in Smith-Magenis syndrome. Am J Med Genet Part C Semin Med Genet. 2010.
15. Faraone S V., Spencer T, Aleardi M, Pagano C, Biederman J. Meta-Analysis of the Efficacy of Methylphenidate for Treating Adult Attention-Deficit/Hyperactivity Disorder. Journal of Clinical Psychopharmacology. 2004.
16. Hanwella R, Senanayake M, de Silva V. Comparative efficacy and acceptability of methylphenidate and atomoxetine in treatment of attention deficit hyperactivity disorder in children and adolescents: A meta-analysis. BMC Psychiatry. 2011.
17. Schachter HM, Pham B, King J, Langford S, Moher D. How efficacious and safe is short-acting methylphenidate for the treatment of attention-deficit disorder in children and adolescents? A meta-analysis. CMAJ. 2001.
18. Cheon KA, Ryu YH, Kim JW, Cho DY. The homozygosity for 10-repeat allele at dopamine transporter gene and dopamine transporter density in Korean children with attention deficit hyperactivity disorder: Relating to treatment response to methylphenidate. Eur Neuropsychopharmacol. 2005.
19. van der Poest Clement E, Jansen FE, Braun KPJ, Peters JM. Update on Drug Management of Refractory Epilepsy in Tuberous Sclerosis Complex. Pediatric Drugs. 2020.
20. De Kuijper G, Evenhuis H, Minderaa RB, Hoekstra PJ. Effects of controlled discontinuation of long-term used antipsychotics for behavioural symptoms in individuals with intellectual disability. J Intellect Disabil Res. 2014.
21. Bradbury J, Avila C, Grace S. Practice-Based Research in Complementary Medicine: Could N-of-1 Trials Become the New Gold Standard? Healthcare. 2020.

22. Schork NJ. Personalized medicine: Time for one-person trials. Nature. 2015.
23. Goetz LH, Schork NJ. Personalized medicine: motivation, challenges, and progress. Fertility and Sterility. 2018.
24. Gupta S, Faughnan ME, Tomlinson GA, Bayoumi AM. A framework for applying unfamiliar trial designs in studies of rare diseases. Journal of Clinical Epidemiology. 2011.
25. Tate RL, Perdices M. Single-Case Experimental Designs for Clinical Research and Neurorehabilitation Settings. Single-Case Experimental Designs for Clinical Research and Neurorehabilitation Settings. 2019.
26. Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D, et al. A clinician's guide for conducting randomized trials in individual patients. CMAJ. 1988.
27. Shamseer L, Sampson M, Bukutu C, Schmid CH, Nikles J, Tate R, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015: Explanation and elaboration. BMJ. 2015.
28. OCEBM Levels of Evidence Working Group, Durieux N, Pasleau F, Howick J. The Oxford 2011 Levels of Evidence. Group. 2011.
29. Porcino AJ, Shamseer L, Chan AW, Kravitz RL, Orkin A, Punja S, et al. SPIRIT extension and elaboration for n-of-1 trials: SPENT 2019 checklist. BMJ. 2020.
30. Mitchell GK, Hardy JR, Nikles CJ, Carmont SAS, Senior HE, Schluter PJ, et al. The effect of methylphenidate on fatigue in advanced cancer: An aggregated N-of-1 trial. J Pain Symptom Manage. 2015.
31. Swanson JM, Volkow ND. Pharmacokinetic and pharmacodynamic properties of stimulants: Implications for the design of new treatments for ADHD. In: Behavioural Brain Research. 2002.
32. Mazefsky CA, Yu L, White SW, Siegel M, Pilkonis PA. The emotion dysregulation inventory: Psychometric properties and item response theory calibration in an autism spectrum disorder sample. Autism Res. 2018.
33. 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.
34. Elliott R, Wagner J, Sales CMD, Rodgers B, Alves P, Café MJ. Psychometrics of the personal questionnaire: A client-generated outcome measure. Psychol Assess. 2016.
35. Goodman R, Meltzer H, Bailey V. The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. International Review of Psychiatry. 2003.
36. Mazefsky CA, Yu L, Pilkonis PA. Psychometric Properties of the Emotion Dysregulation Inventory in a Nationally Representative Sample of Youth. J Clin Child Adolesc Psychol. 2020.
37. Hall CL, Guo B, Valentine AZ, Groom MJ, Daley D, Sayal K, et al. The validity of the Strengths and Difficulties Questionnaire (SDQ) for children with ADHD symptoms. PLoS One. 2019.
38. Kaptein S, Jansen DEMC, Vogels AGC, Reijneveld SA. Mental health problems in children with intellectual disability: Use of the Strengths and Difficulties Questionnaire. J Intellect Disabil Res. 2008.
39. Tate RL, Perdices M, Rosenkoetter U, Wakim D, Godbee K, Togher L, 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.
40. Methylphenidate (MPH): physician's guide to prescribing [Internet]. [cited 2021 Mar 9]. Available from: www.methylphenidate-guide.eu.
41. Senn S. Sample size considerations for n-of-1 trials. Stat Methods Med Res. 2019.
42. Matthijssen AFM, Dietrich A, Bierens M, Kleine Deters R, Van De Loo-Neus GHH, Van Den Hoofdakker BJ, et al. Effects of Discontinuing Methylphenidate on Strengths and Difficulties, Quality of Life and Parenting Stress. J Child Adolesc Psychopharmacol. 2020.
43. Muris P, Meesters C, Van den Berg F. The Strengths and Difficulties Questionnaire (SDQ) further evidence for its reliability and validity in a community sample of Dutch children and adolescents. Eur Child Adolesc Psychiatry. 2003.
44. m-Path [Internet]. [cited 2021 Mar 9]. Available from: www.m-Path.io.

45. Müller AR, Brands MM, van de Ven PM, Roes KC, Cornel MC, van Karnebeek CD, et al. The Power of 1: Systematic Review of N-of-1 Studies in Rare Genetic Neurodevelopmental Disorders. Neurology. 2021.
46. Guyatt GH, Keller JL, Jaeschke R, Rosenbloom D, Adachi JD, Newhouse MT. The n-of-1 randomized controlled trial: Clinical usefulness. Our three-year experience. Ann Intern Med. 1990.
47. Stunnenberg BC, Deinum J, Nijenhuis T, Huysmans F, van der Wilt GJ, van Engelen BGM, et al. N-of-1 Trials: Evidence-Based Clinical Care or Medical Research that Requires IRB Approval? A Practical Flowchart Based on an Ethical Framework. Healthcare. 2020.
48. Kravitz RL, Duan N, Niedzinski EJ, Hay MC, Subramanian SK, Weisner TS. What ever happened to N-of-1 trials? Insiders' perspectives and a look to the future. Milbank Q. 2008.
49. Backman CL, Harris SR. Case studies, single-subject research, and N of 1 randomized trials: Comparisons and contrasts. Am J Phys Med Rehabil. 1999.
50. Gaasterland CMW, Jansen-van der Weide MC, Vroom E, Leeson-Beevers K, Kaatee M, Kaczmarek R, et al. The POWER-tool: Recommendations for involving patient representatives in choosing relevant outcome measures during rare disease clinical trial design. Health Policy (New York). 2018.

Chapter 4

Cannabidiol (Epidyolex®) for severe behavioral manifestations in patients with Tuberous Sclerosis Complex, mucopolysaccharidosis type III and Fragile X syndrome: protocol for a series of randomized, placebo-controlled N-*of*-1 trials

Annelieke R. Müller*, Bibiche den Hollander*, Peter M. van de Ven, Kit C.B. Roes, Lisa Geertjens, Hilgo Bruining, Clara D.M. van Karnebeek, Floor E. Jansen, Marie-Claire Y. de Wit, Leontine W. ten Hoopen, André B. Rietman, Bram Dierckx, Frits A. Wijburg, Erik Boot, Marion M.M.G. Brands, Agnies M. van Eeghen

* Both authors contributed equally.

Submitted

Abstract

Background: Many rare genetic neurodevelopmental disorders (RGNDs) are characterized by intellectual disability (ID), severe cognitive and behavioral impairments, potentially diagnosed as a comorbid autism spectrum disorder or attention-deficit hyperactivity disorder. Quality of life is often impaired due to irritability, aggression and self-injurious behavior, generally refractory to standard therapies. There are indications from previous (case) studies and patient reporting that cannabidiol (CBD) may be an effective treatment for severe behavioral manifestations in RGNDs. However, clear evidence is lacking and interventional research is challenging due to the rarity as well as the heterogeneity within and between disease groups and interindividual differences in treatment response. Our objective is to examine the effectiveness of CBD on severe behavioral manifestations in three RGNDs, including Tuberous Sclerosis Complex (TSC), mucopolysaccharidosis type III (MPS III), and Fragile X syndrome (FXS), using an innovative trial design.

Methods: We aim to conduct placebo-controlled, double-blind, block-randomized, multiple crossover N-*of*-1 studies with oral CBD (twice daily) in 30 patients (aged ≥6 years) with confirmed TSC, MPS III or FXS and severe behavioral manifestations. The treatment is oral CBD up to a maximum of 25 mg/kg/day, twice daily. The primary outcome measure is the subscale irritability of the Aberrant Behavior Checklist. Secondary outcome measures include (personalized) patient-reported outcome measures with regard to behavioral and psychiatric outcomes, disease-specific outcome measures, parental stress, seizure frequency, and adverse effects of CBD. Questionnaires will be completed and study medication will be taken at the participants' natural setting. Individual treatment effects will be determined based on summary statistics. A mixed model analysis will be applied for analyzing the effectiveness of the intervention per disorder and across disorders combining data from the individual N-*of*-1 trials.

Discussion: These N-*of*-1 trials address an unmet medical need and will provide information on the effectiveness of CBD for severe behavioral manifestations in RGNDs, potentially generating generalizable knowledge at an individual-, disorder- and RGND population level.

Trial registration: EudraCT: 2021-003250-23, registered 25 August 2022, https://www.clinicaltrialsregister.eu/ctr-search/trial/2021-003250-23/NL.

Highlights of the study protocol

- Addressing unmet patient needs for treatment of severe behavioral manifestations of rare genetic disorders.
- Evaluation of a novel drug recently approved for certain rare epilepsy syndromes many of which characterized by behavioral problems.
- Innovative trial design (series of N-of-1 trials) in rare disorders, applying evidence-based medicine on an individual as well as group level.
- Use of novel, patient-centered and personalized outcome measures addressing (clinically) relevant items for the patient and caregivers.

Introduction

Rare genetic neurodevelopmental disorders (RGNDs) affect up to 3% of the population.[1] RGNDs are often associated with intellectual disability (ID) and psychiatric comorbidity, including autism spectrum disorder (ASD), and attention-deficit hyperactivity disorder (ADHD), that may result in behavioral manifestations such as irritability, aggression, and self-injurious behavior. Often, these behavioral manifestations pose a great challenge to treating physicians and caregivers as they are refractory to standard psychological, contextual, and pharmacological interventions, and necessitate intensive individual guidance and care. Consequently, these manifestations are associated with the quality of life of not only patients, but also their families, caregivers as well as society. Therefore, there is an urgent unmet need for novel interventional treatment approaches.[2]

Over the last decade, a renewed clinical interest in the use of medicinal cannabis has resulted in promising effects for several indications,[3,4] such as treatment of epilepsy.[5,6] Recently, CBD (Epidyolex®) has been approved by the European Medicines Agency (EMA) to treat refractory epilepsy associated with TSC, Lennox-Gastaut syndrome and Dravet syndrome in patients aged two years and older.[7] There are also indications of efficacy of CBD for severe behavioral manifestations in ID and RGNDs,[8–10] with a favorable side-effect profile compared to currently used medication, such as antipsychotics.[11,12] Additionally, anecdotal reports of families describe a calming effect of medicinal cannabis in some children. As it will be available due to recent market approval, it is important to examine the effectiveness of CBD on behavioral manifestations in RGNDs considering the increasing interest in CBD and urge to treat behavioral problems.

The exact mechanisms of action of CBD are unknown, but previous studies suggest that CBD interacts with many signaling systems, including antagonism of GPR55, desensitization of TRPV1 channels, inhibition of adenosine reuptake, and has neuroprotective and anti-inflammatory effects.[5,6] CBD also affects serotonin 5HT1A signal transduction, gamma-aminobutyric acid (GABA) receptor signaling, and dopamine receptor signaling, processes that are implicated in behavior. Furthermore, CBD is

believed to interact with the endocannabinoid system in several ways,[13] which is involved in regulating a variety of physiological and cognitive processes.[14]

Due to the rarity and heterogeneity of RGNDs, interventional research is challenging. In this study, three RGNDs that are characterized by severe refractory behavioral manifestations and for which CBD has been used at patients' own initiative are Tuberous Sclerosis Complex (TSC), Sanfilippo disease or mucopolysaccharidosis type III (MPS III) and Fragile X syndrome (FXS). These are included as a result of the specific outpatient clinics at the Amsterdam UMC, and due to availability, urgency, and heterogeneity. These patient groups reflect varying neurobiological backgrounds and phenotypes. TSC, MPS III, and FXS are all associated with ID and a severe behavioral phenotype, allowing cross-disorders comparisons. This provides more insight into treatment effects and predictors for treatment response.

TSC is a multisystem, autosomal dominant disorder affecting about 1:6000 children and adults.[15] It is caused by a pathogenic variant in one of two genes, *TSC1* (encoding hamartin) or *TSC2* (encoding tuberin).[16,17] TSC-associated neuropsychiatric disorders (TAND) include epilepsy (85%), ID (50%), ASD (50%), ADHD (30-50%) and behavioral problems (50%).[18,19] Recently, promising results on seizures were found in a randomized, double-blind, controlled trial for CBD in TSC-related seizures in patients with drug-resistant epilepsy, with good efficacy and safety.[20] The use of pharmaceutical-grade CBD in TSC, including relevant mechanism of action, efficacy and safety data, and drug-drug interactions with other anticonvulsant medication was previously described,[21] and a zebrafish model of TSC has been used to examine the influence of CBD on TSC pathology.[22] However, its effect on TSC-related behavioral and cognitive manifestations has not yet been explored sufficiently.[21]

MPS III is an autosomal recessive lysosomal storage disorder caused by a deficiency of 1 of 4 enzymes involved in the degradation of the glycosaminoglycan heparan sulphate. Four types, caused by deficiencies of the different enzymes, are recognized: MPS III type A, B, C and D, with MPS IIIA the most frequent subtype.[23] As a group, MPS III comprises 47%

of all MPS cases diagnosed and the combined birth prevalence is 1.89 per 100.000 live births.[24] It is characterized by progressive intellectual neurologic deterioration (PIND)[25] and severe and progressive behavioral and sleeping problems including restless, destructive, chaotic, anxious and sometimes aggressive behavior.[26] There is yet no curative treatment.[23] To date, no evidence exists for the efficacy of CBD for MPS III, although a potential treatment approach has been described that focuses on modulation of the endocannabinoid system in lysosomal storage disorders including MPS III.[27]

FXS is a relatively more common genetic disorder associated with ID. It is X-linked, occurs in approximately 1:4000 males and 1:8000 females, and is caused by an alteration in the recently renamed Fragile X Messenger Ribonucleoprotein 1 (*FMR1*) gene containing a CGG-repeat with repeat length exceeding 200 CGGs.[28,29] Other manifestations include ADHD (70%), ASD (60%), and anxiety (80%). A recent trial with transdermal CBD gel showed good efficacy on irritability in children with FXS.[9] The role of the endocannabinoid system in FXS, its dysregulation due to the absence of Fragile X Messenger Ribonucleoprotein (FMRP), and the potential role of CBD has been previously described.[13,30] The endocannabinoid system facilitates synaptic homeostasis and plasticity through the cannabinoid receptor 1 (CB1) on presynaptic terminals, resulting in feedback inhibition of neuronal signaling, which are thought to be disrupted in FXS and may be restored by CBD acting as a negative allosteric modulator of CB1.[31] These findings suggest that the endocannabinoid system may be involved in the neurodevelopment and behavior regulation.

Rationale for the N-of-1 design

Trials in RGNDs pose specific methodological challenges due to comorbidities and rarity of conditions.[32,33] Additionally, patients with rare disorders require individualized treatments and outcome measures due to their heterogeneity and vulnerability. It is therefore difficult to conduct traditional randomized controlled trials (RCTs) to determine effectiveness. The N-*of*-1 methodology is an alternative type of RCT, providing rigorous, and the highest level of evidence of treatment effectiveness at an individual level and is consistent with the movement towards personalized care.[34,35] N-*of*-1 studies are randomized, controlled, multiple cross-over trials within

individual patients[36,37] and enhance treatment precision when intervention effects are heterogeneous between individuals.[38,39] In this way, structured and evidence-based treatment decisions can be made for an individual patient. Aggregating the results of several N-*of*-1 trials in different rare, complex, and heterogeneous disorders yields treatment effect estimates[40,41] and contributes to the generalizability to future patients with these RGNDs, but potentially also to patients with other RGNDs.[42]

Aim and objectives

We aim to conduct a series of N-*of*-1 trials to obtain scientific evidence of the effectiveness of treatment with CBD for TSC, MPS III, and FXS. The primary objective is the treatment effect of CBD compared with placebo on irritability. Secondary objectives include assessment of the effect of CBD on psychiatric and behavioral manifestations, disease-specific manifestations, parental stress, seizure frequency, and adverse effects. Personalized outcome measures will be included as well to enable us to take comorbidities into account and to focus on personalized goals. It is hypothesized that CBD has positive effects on severe behavioral manifestations, although interindividual differences in treatment effect might be expected. Baseline characteristics, such as diagnosis, accurate comorbid symptoms, and CYP enzymes enable better interpretation of results and treatment response in these heterogeneous populations with diverse neurobiological and behavioral phenotypes. Thus, a detailed description of the baseline characteristics and demographic information, as well as an extensive set of outcome measures, will provide detailed information about which manifestations may specifically be affected, and help to unravel the mechanism of action of CBD in behavioral manifestations. With that knowledge, CBD may be used as a treatment for other disorders presenting with severe behavioral manifestations. The extensive set of outcome measures will ensure that all essential clinical characteristics of the included patients will be covered. Using a strong methodology, this trial could be considered as both a confirmatory trial for irritability and exploratory for other behavioral and psychiatric outcomes. The current series of trials is part of a project which aims to create more knowledge about the suitability of N-*of*-1 trials and personalized outcome measures for rare disorders in order to facilitate care as well as regulatory decision-making.[43,44]

Methods/Design

Protocol development and patient engagement

The choice of TSC, MPS III, and FXS was based on the severe behavioral manifestations that are an important part of the phenotype and due to our experience with these patient groups. Representatives of the Dutch TSC and FXS patient advocacy organizations, caregivers of patients with TSC and FXS, and clinical experts of all patient groups played an important role in defining knowledge and care gaps, prioritizing the treatment study, selecting outcome measures and developing the current protocol. In the protocol, we have addressed concerns related to caregiver burden and patient burden of participation and issues for recruitment and retention.

The Emma Children's Hospital at the Amsterdam University Medical Center (UMC) is the national expertise center for MPS III. As a result, we have close contact with all Dutch families with MPS III. In addition, we have a longstanding experience in the treatment of behavioral and sleeping problems in these patients. Furthermore, we have a national clinic specialized in TAND at 's Heeren Loo and neuropsychiatric manifestations of FXS, collaborating closely with TSC and epilepsy expertise centers of the UMC Utrecht and expertise center of the University Medical Center of Rotterdam (ENCORE; Genetic NeuroCognitive Developmental Disorders Rotterdam Erasmus Medical Center (MC)).

Study design and duration

We have used the Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) extension for N-*of*-1 trials (SPENT) checklist that is aligned with the CONSORT (consolidated reporting items for trials) extension for N-*of*-1 trials (CENT) for developing this N-*of*-1 protocol.[39]

The study will consist of a series of N-*of*-1 trials followed by an optional open-label extension phase. Each N-*of*-1 trial is block-randomized, placebo-controlled, and double-blinded with multiple crossovers in a single patient. The trial will start with a baseline period of two weeks without any intervention. A variable dose titration phase will follow with a taper period (two weeks) and washout period (one week) before starting the trial. Each

N-*of*-1 trial consists of two cycles, each consisting of one period of CBD treatment (A; six weeks), one period of placebo treatment (B; six weeks), run-in periods (three weeks), taper periods (two weeks), and washouts following A and B (one week) (Figure 1). The total duration of each trial will last around one year. The optional open-label extension phase will be a further twelve months.

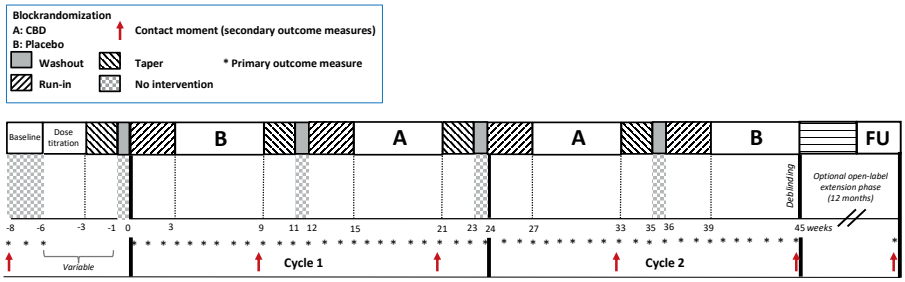


Figure 1. Study design of the N-*of*-trial.

CBD, cannabidiol; FU, follow-up.

Study setting

Prior to the start of the trial, all clinical measures and questionnaires will be completed in the Amsterdam UMC, location AMC. The remaining questionnaires will be completed and study medication will be taken at the participants' natural setting.

Recruitment

Participants will be recruited through the outpatient clinics for TSC and FXS at 's Heeren Loo, the UMC Utrecht and the Erasmus MC, and through patient organizations and the MPS III expert center at the Amsterdam UMC.

Study population

The study population consists of children and adults with TSC and FXS, and children with MPS III, all suffering from severe behavioral manifestations. We aim to conduct a patient-centered trial, allowing for a natural setting and flexibility, including continuation of concurrent therapies.

Inclusion criteria include:

- Minimum age of 6 years old.
- Clinically and/or genetically definite diagnosis of TSC, MPS III or FXS (modified Gomez Criteria, clinical criteria, positive genetic test, or enzyme deficiency).
- Suffering from severe behavioral manifestation with a minimum score of 4 on the Clinical Global Impression - severity (CGI-S) scale.[45]
- Stable dose of all psychopharmacological medications or interventions for one month prior to screening and willingness of the participant and legal representatives to maintain the current medication regimen throughout the trial.
- Presence of a consistently available patient caregiver for proxy reports.

Exclusion criteria include:

- Any known or suspected hypersensitivity to cannabinoids or any of the excipients of the Investigational Medicinal Product (IMP), such as sesame oil.
- History of recreational or medicinal cannabis, or cannabinoid-based medications, with three months prior to screening and the patient is unwilling to abstain for the duration of the study.
- History or current evidence of significantly impaired liver function, defined as 1) Alanine aminotransferase (ALT) and/or aspartate aminotransferase (AST) > 5 x upper limit of normal (ULN); 2) ALT or AST > 3 x ULN with concomitant total bilirubin > 2.0 x ULN; or 3) ALT or AST ≥ 3 x ULN with the appearance of fatigue, nausea, vomiting, right upper quadrant pain or tenderness, fever, rash, and/or eosinophilia.
- Pregnancy or breastfeeding. Females of childbearing potential must be willing to use an effective method of contraception from the time of consent until six weeks after treatment discontinuation and inform the trial if pregnancy occurs.
- Glaucoma.
- History of general anesthesia in the four weeks prior to enrolment.

- Use of any interfering medication within 30 days prior to enrolment or planning to take interfering medication during the trial.
- Any planned major surgery within the duration of the trial.
- Expected inability to undergo blood sampling due to anxiety or resistance.
- Unwillingness or inability to swallow the study drug (or placebo).

In addition to the exclusion criteria, use of valproate should be stable three months prior to enrolment.

Sample size

The sample size calculation was based on a summary measures analysis of the treatment effect.[46] The difference between the mean irritability ratings of the Aberrant Behavior Checklist (ABC) in CBD periods and placebo periods was used as a summary measure for the treatment effect in an individual participant. Heussler et al. (2019) reported a standard deviation (SD) of 12 points for single ratings on the subscale.[9] This estimate for the SD includes both within- and between-subject variance. The intraclass correlation coefficient (ICC) was assumed to be 0.75. In addition to the estimate for the within-subject variation of the outcome measure, an a priori estimate was needed for the between-subject variation of the treatment effect. Assuming an SD of 3 points for the SD of the random treatment effect, 95% of the subject-specific treatment effects roughly falls within a range of 12 points. Based on the estimate assuming two cycles with six ABC ratings within each period, a total of 6 participants will yield 80% power to detect a mean difference of 6 points between intervention and placebo periods when assuming a two-sided significance level of 5%. As effects for pediatric and adult population may differ, separate power analyses were performed for the pediatric and adult cohorts. Per cohort, this amounts to 6 pediatric patients and/or 6 adult patients, with a total inclusion of 12 patients with TSC, 12 with FXS and 6 pediatric patients with MPS III. MPS III is a type of childhood dementia and most patients never reach adulthood.[26] Adult patients experience progressive dementia. This stage of the disease is not associated with severe behavioral manifestations, and therefore we will not include adult patients with MPS III.

Blinding, treatment allocation, randomization

Participants, parents, caregivers, physicians and researchers will all be blinded during the trial. The random allocation sequence will be generated for block randomization in a 1:1 ratio and implemented by the hospital pharmacist and sequentially numbered packages. Unblinding will occur when a participant has completed the two cycles or in case of a serious adverse event (SAE) that cannot be treated without knowing which treatment the patient was receiving. Investigators involved in data analysis will remain blinded until the end of the follow-up period.

Interventions and dosing schedule

Patients will receive a pharmaceutical formulation of highly purified CBD derived from *Cannabis sativa* L. (100 mg/mL) oral solution (Epidyolex® [Jazz Pharmaceuticals]) alternately with a placebo distributed by the Amsterdam UMC hospital pharmacist. CBD reduced TSC-associated seizures versus placebo with similar efficacy between the 25 and 50 mg/kg/d doses.[20] Given that the safety profile of 25 mg/kg/d was superior to 50 mg/kg/d, the lower dose range suggests a superior benefit-to-risk ratio. Standard rules for the use of CBD are in force. Participants can continue their psychopharmacological medications.

Dose titration phase

Prior to the N-*of*-1 trials and following the baseline period, a dose titration phase will take place, comprising escalating doses of CBD with twice daily administration from 2.5 mg/kg/day up to 25 mg/kg/day. Dose escalation steps involve an increase of 2.5 mg/kg/day. Adverse effects during the dose titration phase will be checked twice a week by a video or phone call. Also, hepatic enzyme levels will be measured at baseline and weekly from the third week of the titration phase, unless indication requires deviation. In case of adverse effects, or if the hepatic enzyme levels are ≥2 higher than the levels measured at baseline, the lower dose (2.5 mg/kg lower) will be taken. The final dosage will be based on the dose titration phase, with the highest dosage applied during this phase with the least adverse effects. The length of the dose titration phase will vary depending on the final dosage, followed by a taper and washout period.

Rescue medication

The study protocol does not specify any rescue medication to treat side effects of CBD treatment because, based on previous studies on the safety of CBD, these are usually mild at the doses used in this study. In case of SAEs, the patient will be withdrawn from the trial and appropriate treatment will be started.

Outcome measures

The primary outcome is a change on the irritability subscale score of the Aberrant Behavior Checklist (ABC-I) during active interventional periods compared to placebo periods.

Secondary outcome measures include:

- Total ABC;[47]
- CGI;[45]
- Syndrome-specific outcome measures, including the TSC-specific patient-reported outcome measure (TSC-PROM)[48] and the Sanfilippo Behavior Rating Scale (SBRS);[49]
- Pediatric Quality of Life Inventory (PedsQL);[50]
- Anxiety, Depression, and Mood Scale (ADAMS);[51]
- Social Communication Questionnaire (SCQ);[52]
- Social Responsiveness Scale (SRS)-2 (when appropriate);[53]
- Short Sensory Profile (SSP-NL);[54]
- Parenting Stress Questionnaire (OBVL);[55]
- Goal Attainment Scaling (GAS);[56]
- Personal Questionnaire (PQ);[57]
- Focal and generalized seizure frequency;
- Adverse effects;
- Hepatic enzyme levels.

Baseline clinical characteristics, demographic information, medical history, results on CYP enzymes, and information regarding diagnosis will be collected and recorded in detail to enable better interpretation of results and explore factors associated with treatment response, because of the heterogeneity of the population and diverse neurobiological and behavioral phenotypes.[44] Collecting data on CYP enzymes has already been

best practice and recommended to measure for refractory or difficult to treat behavioral manifestations in this population. The shortened version Vineland Adaptive Behavior Scales-III (VABS-III)[58] will be filled out by the clinicians to determine whether the SRS-2 could be filled out during the trial.[59] Optionally, electroencephalogram (EEG) recordings will be performed at baseline to detect changes in brain resting-state EEG for stratification of responsiveness post hoc.

Rationale for outcome measures

As a primary outcome measure, the ABC-I was selected. The ABC is a caregiver-completed rating scale for assessing problem behaviors in children and adults, which has robust psychometric properties in intellectually impaired and developmentally delayed populations.[47,60,61] The empirically derived and widely used irritability subscale consists of 15 items and comprises items reflecting temper outbursts, aggression, negative affect, and self-harm behavior. Irritability has been identified as a prominent behavioral correlate of ASD. The ABC-I has often been used as an outcome measure in treatment studies of behavioral problems in individuals with ASD and ID.[62–64] The ABC, including the irritably subscale, has been shown to be sensitive to treatment change in previous clinical trials of FXS.[65,66] The other domains of the ABC, including lethargy/social withdrawal, stereotypic behavior, hyperactivity/noncompliance, and inappropriate speech, will serve as generalization measures to evaluate transfer effects of the intervention to a broader domain of functioning.[67] A generalization measure is an outcome measure that is related to the target behavior (irritability), for example irritability at school and at home (another setting), or interventional effects on a completely different behavior, such as less hyperactive when the target behavior is irritability.

By including an extensive set of secondary outcome measures in this study, we aim to explore the effectiveness of CBD on several behavioral and psychiatric domains as it is yet unclear which manifestations specifically respond to treatment. Moreover, we aim to explore if these outcome measures are appropriate and useful in these patient groups and in RGNDs in general. We chose outcome measures at different levels, such as disorder-specific, personalized and generalizable measures. As most

of our patients will have ID, proxy-rated outcome measures applicable to children as well as adults were selected that have been psychometrically considered valid tools to measure aberrant behavior, anxiety, mood, ASD features, and parental stress in ID.

The CGI scale is a well-established rating tool applicable to all psychiatric disorders that can easily be used by the practicing clinician and provides an assessment of the clinician's view of the patient's global functioning prior to and after initiating a study medication.[45] The CGI has two components: the CGI-Severity (CGI-S) which rates illness severity, and the CGI-Improvement (CGI-I) which rates change from the initiation (baseline) of treatment. The CGI can track clinical progress across time and has shown to correlate well with standard, well-known research drug efficacy scales and longer, more tedious and time consuming rating instruments across a wide range of psychiatric diagnoses.[68,69] However, the CGI is not goal-oriented and changes do not provide mechanistic insight.

Additionally, available syndrome-specific outcome measures will be included for both TSC (TSC-PROM) and MPS III (SBRS). These PROMs focus on syndrome-specific targets that are of relevance to these patients and their functioning and quality of life. A recent validation study of the PROM for adults with TSC has shown that it is a reliable and valid instrument to measure the impact of the disease on functioning, which can be used in clinical and research settings to systematically gain insight into patients' experiences.[48] It covers the physical, mental health domain, activities and participation and environmental factors, addressing the impact of specific TSC manifestations on adult patients' health-related quality of life (QoL). The SBRS is developed to assess the behavioral phenotype in children with MPS III and its progression and results from treatment over time.[49] The SBRS is validated in 25 children with MPS IIIA, aged 2 to 18 years old.[70] As there is no specific questionnaire on QoL for FXS and children with TSC, the PedsQL will be used, which is a practical, brief, standardized, generic assessment tool to measure health-related QoL. Next to the pediatric version, an adults version exists which will be used for adults.[50]

The ADAMS, considered a psychometrically sound instrument among individuals with ID,[51] will be used to screen for symptoms of anxiety, depression and mood disorders.

The SCQ (originally the Autism Screening Questionnaire (ASQ)) will be used to assess the severity of ASD symptoms.[71–74] The SCQ current version will be used as it enables us to screen for ASD, to compare overall levels or severity of ASD symptoms, and to assess current ASD symptoms and change over time in both young and older children, adolescents, and adults.

The SRS-2 is a widely used measure of ASD symptoms for social behavioral problems in children and adults.[53] It may not be suitable for patients with a developmental age below four years. It will therefore be filled out when applicable, depending on the developmental age as assessed by the VABS-III during baseline.[59]

The SSP is a commonly used shortened form of Dunn's Sensory Profile caregiver questionnaire,[75] containing 38 items measuring sensory features, organized into seven subscales. The total score will be used to measure sensory functioning.[76]

Reducing severe behavioral manifestations may have benefits for family and caregivers. Therefore, we chose the parenting stress questionnaire (OBVL) which is applicable to children of all ages and has been validated for institutions for youth care, including mild ID as well.[55]

Furthermore, GAS and the PQ enable us to focus on personalized and for participants relevant targets, also reflecting the treatment target.[56,77] GAS is an individualized outcome measure, involving goal selection and scaling standardized to calculate the extent to which an individual's goal is met. Patients and caretakers will select their own specific goals together with their treating physician/therapist. It is a measurement instrument that is very sensitive to change, in particular in small and heterogeneous patient populations.[56] The PQ is used as a symptom list to compare assessments of personalized goals with those measured by GAS.[57] Adding the PQ allows us to compare outcomes of standardized and personalized tools.

In case participants have epilepsy, a seizure diary will be used to evaluate change in seizure frequency.

Hepatic enzyme levels (ALT, AST, and bilirubin) will be checked to monitor adverse effects.

Trial procedures

During the N-*of*-1 trial, the final CBD dosage as determined by the dose titration phase or placebo will be administered twice daily. During washout periods, no study medication will be taken.

N-of-1 trial: multiple crossover phase

Prior to the start of the trials, the participant and substitute decision maker(s) will be seen at the clinic to discuss the procedure and sign for informed consent (Figure 2). Personalized goals with regard to GAS will be identified together with the treating physician, psychologist, patients and primary caregivers. Seizures semiology will be discussed in detail and classified according to the classification of the International League Against Epilepsy (ILAE).[78] Reporting of seizure frequency (by patients or caregivers) will be assessed. The shortened version VABS-III will be filled out by the clinicians to determine whether the SRS-2 could be filled out during the trial.[59] Optionally, EEG recordings will be performed at baseline.

A two-week baseline period will follow without any intervention. The ABC, CGI, syndrome-specific outcome measures, ADAMS, SCQ, SRS-2 (if applicable), SSP-NL, and OBVL, will be filled out and seizures will be reported. These questionnaires can be completed within one hour. For children, the OBVL will be filled out by parents or primary caregivers if they have known the child for at least six months. The CGI can be completed in less than a minute by an experienced rater. A dose titration phase is followed by a taper period and washout period. During the dose titration phase, contact moments will take place (by phone) twice weekly, and hepatic enzyme levels will be measured. The individual N-*of*-1 trial consists of two cycles each containing one active treatment (A), one placebo treatment (B), run-in periods, taper periods, and washout periods. Medication will be administered at home, institution setting or day care by caregivers. The ABC irritability subscale will

be filled out weekly by primary caregivers, which can be completed within a few minutes. The other outcome measures will be scored once during the baseline period, at the end of each interventional (including placebo) period, with a total of five measurements, and if participating in the optional extension phase at the follow-up measurement. Adverse effects will be assessed as well. The questionnaires can be filled out digitally (Castor EDC), phone or on paper forms. Caregivers will be asked to report all seizures in the seizure diary. To reduce burden, assessments occur by phone calls except for the clinical visits. During the experiment, it will be attempted to not switch rating caregivers.

Optional open-label extension phase

In consultation with the primary caregivers, patients may continue with CBD treatment during an optional one-year open-label extension phase after which a final contact moment takes place. At this follow-up measurement, questionnaires will be filled out again and personalized goals will be evaluated.

Safety evaluation

Adverse events

Adverse events (AEs) will be monitored throughout the N-*of*-1 trials. All SAEs and Suspected Unexpected Serious Adverse Reactions (SUSARs) occurring during the study and either reported spontaneously or as part of the AE monitoring will be followed up by the principal investigator and will be reported separately to the Medical Ethics Committee (MEC). Unblinding will occur if there is reason to believe a SAE or SUSAR was due to the study medication and if the patient cannot be treated without knowing which treatment they were receiving. Unblinding and the reason for unblinding will be recorded. AEs, SAEs and SUSARs will be followed until they have abated or a stable situation has been reached.

Removal from the trial and replacement of participants

Participants will be removed from the study if informed consent is withdrawn. The investigator can decide to withdraw a participant from the N-*of*-1 trial for urgent medical reasons. Participants with hepatic enzyme level elevations sex times or greater the levels measured during baseline will be excluded, as this is a known potential for drug-induced liver injury with CBD.

Time point (weeks; variable)		Enrollment*	Allocation	Baseline	Dose titration	Taper	Washout	Cycle 1	Cycle 2	Follow-up
				0	2	5	7	8-31	32-55	
Enrolment	Eligibility screen	X								
	Informed consent	X								
	Allocation		X							
Interventions	CBD				X	X		X	X	X
	Placebo							X	X	X
Assessments	VABS-III	X								
	EEG recordings	X								
	ABC irritability (weekly)			X				X	X	X
	Total ABC			X				X	X	X
	CGI			X				X	X	X
	Syndrome-specific outcome measures / PedsQL			X				X	X	X
	ADAMS			X				X	X	X
	SCQ			X				X	X	X
	SRS-2			X				X	X	X
	SSP			X				X	X	X
	OBVL			X				X	X	X
	GAS	X		X				X	X	X
	PQ	X		X				X	X	X
	Seizure frequency			X	X	X	X	X	X	X
	Hepatic enzyme levels			X	X			X	X	X
	Side effects				X	X		X	X	X

Figure 2. Time schedule of enrolment, interventions, and assessments.

Crosses (X) indicate research steps to be conducted. Underlined crosses (X) indicate assessments via phone calls. ABC, Aberrant Behavior Checklist; ADAMS, Anxiety, Depression, and Mood Scale; CBD, cannabidiol; CGI, Clinical Global Impression; EEG, electroencephalogram; GAS, Goal Attainment Scaling; OBVL, parenting stress questionnaire; PedsQL, Pediatric Quality of Life Inventory; PQ, personal questionnaire; SCQ, Social Communication Questionnaire; SSP, Short Sensory Profile; SRS-2, Social Responsiveness Scale; VABS-III, Vineland Adaptive Behavior Scale-III.

In case of a drop-out, completed weeks before withdrawal will still be analyzed and a new participant will be recruited with a newly randomized sequence.

Premature termination of the study

The study will be terminated prematurely if more than one participant indicates a great burden of switching between placebo and CBD periods accompanied by serious safety concerns.

Data safety monitoring board (DSMB) and monitoring

Independent qualified monitors from the Clinical Monitoring Center (CMC) will monitor the study. Because of the small sample size and the close monitoring of patients, it was not necessary to set up a DSMB.

Ethics approval and consent to participate

This protocol is approved by the institutional review board of the Amsterdam UMC, location AMC (2022_0580). The study will be conducted in accordance with the principles of the Declaration of Helsinki (2013), the EU General Data Protection Regulation, the Dutch Medical Research Involving Human Subjects Act and Good Clinical Practice guidelines, the Dutch Act on Implementation of the General Data Protection Regulation, and the Amsterdam UMC standard operating procedures. The study is registered under EudraCT number 2021-003250-23, protocol no. NL78162.018.21, version 4.0, registration date 25 August 2022.

Participants and caregivers will receive a letter with information about the objectives, background and study procedures with a maximum of three months to consider their decision regarding the study's progress. Given that the study population includes children, and as most of the participants will suffer from ID, the legal representative(s) will be asked to sign the informed consent. Participants will be informed about the study adjusted to their level of intellectual functioning. Participants and legal representatives are allowed to contact the research team when remaining questions are present during the entire trial period.

Analysis

Data collection and management

All data will be collected and handled in accordance with applicable privacy and data protection regulations. We will collect the data according to the FAIR principles (Findable, Accessible, Interoperable, Reusable).[79] Trial-

specific documents and the Case Report Forms (CRFs) will be securely stored with restricted access limited to nominated research staff.

Assessments will be entered in the CRFs set up in Castor Electronic Data Capture (EDC). All questionnaires can be filled out digitally (Castor EDC) or by using paper forms. Automatic reminders will be sent when a questionnaire has not been completed on time. Participant burden will be limited as much as possible by using a subscale of the ABC for weekly questionnaires and having contact moments by video-conference or phone instead of a visit.

A participant identification code list will be used with unique participant identifiers not deducible to participants. Only four investigators and a methodologist and biostatistician will have access to the key and source data. Data will be stored for 25 years in a secured database and body material will be stored for 5 years.

Statistical methods

Individual treatment effects for participants will be quantified as summary statistics. A mixed model analysis will be applied for statistically testing the effectiveness of the intervention at group level combining data from the individual N-*of*-1 trials.[80] The mixed model will account for between-subjects heterogeneity in interventional effects by including a random treatment effect.

The mean treatment effect on the primary outcome irritability subscale of the ABC will be estimated and tested using a linear mixed model. The model will contain a fixed effect for the average treatment effect (CBD or placebo), random effects for patients, cycle within patient, and treatment (within patient). A mixed model analysis with similar model structure will be performed for the secondary study parameters. Because of the many data points per period, small amounts of missing data will not pose problems for the mixed model analysis, assuming missingness is random. The lmer package of R will be used for mixed model analyses. An analysis based on a summary measure will be performed if issues such as singularity arise. A two-sided significance level of 5% will be used.

EEG data will be processed offline using the Neurophysiological Biomarker Toolbox,[81] an open-source MATLAB toolbox for the computation and integration of neurophysiological biomarkers. With this toolbox, a wide array of resting state parameters can be evaluated including power spectra and coherence and has been applied to neurological clinical studies. The data processing will be performed using MATLAB 7.12.0 software.[82]

Discussion

With the proposed series of randomized, double-blind N-*of*-1 trials with open-label extension phase, the effectiveness of CBD on severe behavioral manifestations in TSC, MPS III and FXS will be evaluated. TSC, MPS III and FXS are distinct RGNDs with unique clinical features. They share similarities in terms of their neurological involvement, ID, and behavioral challenges. However, they differ in terms of the underlying genetic variants, disease mechanisms, physical manifestations, and prevalence. It is crucial to note that each condition exhibits a wide range of symptoms and severity. Using a single study protocol for multiple disorders offers advantages such as increased efficiency, larger samples sizes, and comparative analysis opportunities, and can be considered a basket trial.[83,84] Especially in rare, complex, and heterogeneous disorders such as these, series of N-*of*-1 trials enable determination of the treatment effects in individual patients as well as at the group level. In this way, structured and evidence-based treatment decisions can be made for an individual patient at risk for trial and error approach, and cross-over disease comparison together with medical, in-depth and mechanistic information will produce generalizable knowledge that can be applied to future patients with RGNDs.

N-*of*-1 studies are recommended in rare genetic disorders when the intervention has a predictable duration of effect and low recruitment rate is expected, like this proposed trial.[44] An explanation about the suitability of N-*of*-1 studies in RGNDs in terms of heterogeneity, personalization, design, outcome measures, and the analyses was provided in a recently published N-*of*-1 study protocol.[44] The N-*of*-1 approach to estimating population effects may come with a caveat regarding generalizability of the results. To tackle

this challenge, we included three disease groups in our study. Combining the results of the N-*of*-1 trials in different patient groups potentially yields information that may be extrapolated to the RGNDs population level.[42]

CBD is expensive, costing about 500 US dollars per month for rare epilepsies, although highly depending on dosages. If this study shows efficacy of CBD for severe behavioral manifestations, our major goal is to get CBD accessible to those patients who are expected to benefit, which could be facilitated by licensing and reimbursement by healthcare authorities and insurances. Before the start of the trial, we consulted ZIN (The National Health Care Institute in the Netherlands, "Zorginstituut Nederland") and CBG (the Dutch Medicines Evaluation Board, "College ter Beoordeling van Geneesmiddelen") on the study design, inclusion and exclusion criteria and outcome measures. Advises on outcome measures included justification for choice of the ABC-I subscale, particularly as this study aims for a broad indication of behavioral problems not associated with a specific syndrome, and defining and justifying a testing hierarchy for the endpoints with the CGI high in the testing hierarchy. An extensive list of secondary endpoints is generally accepted for an exploratory study, while limiting the number of endpoints to those relevant to support the claimed indication is recommended for a confirmatory study. Furthermore, several approaches were suggested for supporting extrapolation, such as substantiation through mechanism of action with similar effects of CBD on behavioral manifestations regardless of the neurodevelopmental disorder, subgroup analysis per disorder indicating the absence of effect modification, and inclusion of additional neurodevelopmental disorders.

The burden for the patient of the study is mostly caused by the use of blinded cross-over periods, the use of placebo, the prolonged dose titration phase, and the filling in of questionnaires by caregivers. The benefits of the study include the fact that patient-centered N-*of*-1 studies may help individuals to better self-manage their behavioral symptoms. The patients involved in the N-*of*-1 trials may draw immediate benefit from the trial as every patient is exposed to the treatment with CBD, and the N-*of*-1 design will enable an individual treatment decision in terms of evidence-based medicine. This is unlike many population-based trials where, depending on the protocol

and design used, an individual may have been on a placebo for the entire trial. Moreover, data from the current series of N-*of*-1 trials will be pooled to obtain a population treatment effect estimate.

In conclusion, we consider that the N-*of*-1 trial design is excellent to study pharmacological treatments of disease manifestations in rare populations. The current study will provide crucial information about the efficacy of CBD for severe behavioral manifestations in these complex and vulnerable patient populations.

Acknowledgements
The authors would like to thank the Dutch patient advocacy organizations (Stichting Tubereuze Sclerosis Nederland, Fragiele X Vereniging Nederland) for their feedback on the study protocol. The Amsterdam UMC and Erasmus MC are member of European Reference Network (ERN) ITHACA. The Amsterdam UMC is member of MetabERN. The UMC Utrecht is member of ERN EpiCare.

Funding
The trial is financially sponsored by a grant from GWResearch/Jazz pharmaceuticals and the Dutch TSC foundation. The open-label extension is also funded by Jazz Pharmaceuticals.

Competing interests
The authors declare that they have no competing interest.

Authors' contributions
AM and AvE proposed the study. AM and BdH initiated the design and wrote the study protocol. PvdV designed the statistical analysis and performed the analyses. KR provided methodological expertise. LG conducted EEG measurements. HB supervised the study design and conducted EEG measurements. CvK, FJ, MdW, LtH, AR, BD, FW, and EB supervised the study design. MB and AvE supervised the study design and wrote the study protocol. All authors provided critical feedback, read, and approved the final manuscript.

References

1. Woittiez, I., Putman, L., Eggink, E., Ras, M., & Sterkenburg PS. Zorg Beter Begrepen. Soc en Cult Planbureau [Internet]. 2014; Available from: http://pdf.swphost.com/Sozio/e-nieuwsbrief/2014/Zorg_beter_begrepen_drukversie.pdf.
2. Hunt A. Tuberous sclerosis: a survey of 97 cases. Dev Med Child Neurol. 1983;
3. Iseger TA, Bossong MG. A systematic review of the antipsychotic properties of cannabidiol in humans. Schizophrenia Research. 2015.
4. MacCallum CA, Russo EB. Practical considerations in medical cannabis administration and dosing. European Journal of Internal Medicine. 2018.
5. Devinsky O, Cilio MR, Cross H, Fernandez-Ruiz J, French J, Hill C, et al. Cannabidiol: Pharmacology and potential therapeutic role in epilepsy and other neuropsychiatric disorders. Epilepsia. 2014.
6. Franco V, Perucca E. Pharmacological and Therapeutic Properties of Cannabidiol for Epilepsy. Drugs. 2019.
7. GW Pharma Ltd. Epidyolex (cannabidiol) oral solution [Internet]. [cited 2023 May 9]. Available from: https://www.medicines.org.uk/ emc/product/10781.
8. Tartaglia N, Bonn-Miller M, Hagerman R. Treatment of Fragile X Syndrome with Cannabidiol: A Case Series Study and Brief Review of the Literature. Cannabis and Cannabinoid Research. 2019.
9. Heussler H, Cohen J, Silove N, Tich N, Bonn-Miller MO, Du W, et al. A phase 1/2, open-label assessment of the safety, tolerability, and efficacy of transdermal cannabidiol (ZYN002) for the treatment of pediatric fragile X syndrome. J Neurodev Disord. 2019.
10. Efron D, Taylor K, Payne JM, Freeman JL, Cranswick N, Mulraney M, et al. Does cannabidiol reduce severe behavioural problems in children with intellectual disability? Study protocol for a pilot single-site phase I/II randomised placebo controlled trial. BMJ Open. 2020.
11. McLaren JL, Lichtenstein JD, Metcalfe JD, Charlot LR, Drake RE, Beasley JB. Psychotropic Use Among Youths With Intellectual and Developmental Disabilities. Psychiatr Serv. 2021.
12. Deb S, Nancarrow T, Limbu B, Sheehan R, Wilcock M, Branford D, et al. UK psychiatrists' experience of withdrawal of antipsychotics prescribed for challenging behaviours in adults with intellectual disabilities and/or autism. BJPsych Open. 2020;6(5).
13. Palumbo JM, Thomas BF, Budimirovic D, Siegel S, Tassone F, Hagerman R, et al. Role of the endocannabinoid system in fragile X syndrome: potential mechanisms for benefit from cannabidiol treatment. J Neurodev Disord. 2023;15(1):1–10.
14. Zou S, Kumar U. Cannabinoid receptors and the endocannabinoid system: Signaling and function in the central nervous system. Vol. 19, International Journal of Molecular Sciences. 2018.
15. Webb DW, Osborne JP. Tuberous sclerosis. Archives of Disease in Childhood. 1995.
16. Crino PB, Nathanson KL, Henske EP. Medical progress: The tuberous sclerosis complex. N Engl J Med. 2006;
17. Young J, Povey S. The genetic basis of tuberous sclerosis. Molecular Medicine Today. 1998.
18. de Vries PJ, Wilde L, de Vries MC, Moavero R, Pearson DA, Curatolo P. A clinical update on tuberous sclerosis complex-associated neuropsychiatric disorders (TAND). Vol. 178, American Journal of Medical Genetics, Part C: Seminars in Medical Genetics. 2018.
19. Curatolo P, Moavero R, de Vries PJ. Neurological and neuropsychiatric aspects of tuberous sclerosis complex. The Lancet Neurology. 2015.
20. Thiele EA, Bebin EM, Bhathal H, Jansen FE, Kotulska K, Lawson JA, et al. Add-On Cannabidiol Treatment for Drug-Resistant Seizures in Tuberous Sclerosis Complex: A Placebo-Controlled Randomized Clinical Trial. JAMA Neurol. 2020.
21. Samanta D. A scoping review on cannabidiol therapy in tuberous sclerosis: Current evidence and perspectives for future development. Vol. 128, Epilepsy and Behavior. 2022.

22. Serra I, Scheldeman C, Bazelot M, Whalley BJ, Dallas ML, de Witte PAM, et al. Cannabidiol modulates phosphorylated rpS6 signalling in a zebrafish model of Tuberous Sclerosis Complex. Behav Brain Res. 2019;
23. Valstar MJ, Neijs S, Bruggenwirth HT, Olmer R, Ruijter GJG, Wevers RA, et al. Mucopolysaccharidosis type IIIA: Clinical spectrum and genotype-phenotype correlations. Ann Neurol. 2010.
24. Poorthuis BJHM, Wevers RA, Kleijer WJ, Groener JEM, De Jong JGN, Van Weely S, et al. The frequency of lysosomal storage diseases in The Netherlands. Hum Genet. 1999;105(1–2).
25. Warmerdam HAG, Termeulen-Ferreira EA, Tseng LA, Lee JY, van Eeghen AM, Ferreira CR, et al. A Scoping Review of Inborn Errors of Metabolism Causing Progressive Intellectual and Neurologic Deterioration (PIND). Front Neurol. 2020;10.
26. Valstar MJ, Ruijter GJG, van Diggelen OP, Poorthuis BJ, Wijburg FA. Sanfilippo syndrome: A mini-review. Journal of Inherited Metabolic Disease. 2008.
27. Schuchman EH, Ledesma MD, Simonaro CM. New paradigms for the treatment of lysosomal storage diseases: targeting the endocannabinoid system as a therapeutic strategy. Vol. 16, Orphanet Journal of Rare Diseases. 2021.
28. Hagerman RJ, Berry-Kravis E, Hazlett HC, Bailey DB, Moine H, Kooy RF, et al. Fragile X syndrome. Vol. 3, Nature reviews. Disease primers. 2017.
29. Khandjian EW, Robert C, Davidovic L. FMRP, a multifunctional RNA-binding protein in quest of a new identity. Front Genet. 2022;13:2035.
30. Jung KM, Sepers M, Henstridge CM, Lassalle O, Neuhofer D, Martin H, et al. Uncoupling of the endocannabinoid signalling complex in a mouse model of fragile X syndrome. Nat Commun. 2012;3.
31. Laprairie RB, Bagher AM, Kelly MEM, Denovan-Wright EM. Cannabidiol is a negative allosteric modulator of the cannabinoid CB1 receptor. Br J Pharmacol. 2015;172(20).
32. Gupta S, Faughnan ME, Tomlinson GA, Bayoumi AM. A framework for applying unfamiliar trial designs in studies of rare diseases. Journal of Clinical Epidemiology. 2011.
33. Tate RL, Perdices M. Single-Case Experimental Designs for Clinical Research and Neurorehabilitation Settings. Single-Case Experimental Designs for Clinical Research and Neurorehabilitation Settings. 2019.
34. Bradbury J, Avila C, Grace S. Practice-Based Research in Complementary Medicine: Could N-of-1 Trials Become the New Gold Standard? Healthcare. 2020.
35. OCEBM Levels of Evidence Working Group, Jeremy Howick, Iain Chalmers (James Lind Library), Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, Bob Phillips, Hazel Thornton OG and MH. The Oxford levels of Evidence 2. Oxford Cent Evidence-Based Med. 2011.
36. Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D, et al. A clinician's guide for conducting randomized trials in individual patients. CMAJ. 1988.
37. Shamseer L, Sampson M, Bukutu C, Schmid CH, Nikles J, Tate R, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015: Explanation and elaboration. BMJ. 2015.
38. OCEBM Levels of Evidence Working Group, Durieux N, Pasleau F, Howick J. The Oxford 2011 Levels of Evidence. Group. 2011.
39. Porcino AJ, Shamseer L, Chan AW, Kravitz RL, Orkin A, Punja S, et al. SPIRIT extension and elaboration for n-of-1 trials: SPENT 2019 checklist. BMJ. 2020.
40. Stunnenberg BC, Raaphorst J, Groenewoud HM, Statland JM, Griggs RC, Woertman W, et al. Effect of Mexiletine on Muscle Stiffness in Patients with Nondystrophic Myotonia Evaluated Using Aggregated N-of-1 Trials. JAMA - J Am Med Assoc. 2018;320(22).
41. 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.
42. Mitchell GK, Hardy JR, Nikles CJ, Carmont SAS, Senior HE, Schluter PJ, et al. The effect of methylphenidate on fatigue in advanced cancer: An aggregated N-of-1 trial. J Pain Symptom Manage. 2015.

43. Müller AR, Zinkstok JR, Rommelse NNJ, van de Ven PM, Roes KCB, Wijburg FA, et al. Methylphenidate for attention-deficit/hyperactivity disorder in patients with Smith–Magenis syndrome: protocol for a series of N-of-1 trials. Orphanet J Rare Dis. 2021;16(1).
44. Müller AR, Brands MM, van de Ven PM, Roes KC, Cornel MC, van Karnebeek CD, et al. The Power of 1: Systematic Review of N-of-1 Studies in Rare Genetic Neurodevelopmental Disorders. Neurology. 2021.
45. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. Psychiatry (Edgmont). 2007;4(7).
46. Senn S. Sample size considerations for n-of-1 trials. Stat Methods Med Res. 2019.
47. Aman MG, Singh NN, Stewart AW, Field CJ. Psychometric characteristics of the aberrant behavior checklist. Am J Ment Defic. 1985.
48. Müller AR, Luijten MAJ, Haverman L, De Ranitz-Greven WL, Janssens P, Rietman AB, et al. Understanding the impact of Tuberous Sclerosis Complex: Development and validation of the TSC-PROM. Press. 2023.
49. Shapiro EG, Escolar ML, Delaney KA, Mitchell JJ. Assessments of neurocognitive and behavioral function in the mucopolysaccharidoses. Molecular Genetics and Metabolism. 2017.
50. Varni JW, Seid M, Rode CA. The PedsQL™: Measurement model for the pediatric quality of life inventory. Med Care. 1999;37(2).
51. Esbensen AJ, Rojahn J, Aman MG, Ruedrich S. Reliability and Validity of an Assessment Instrument for Anxiety, Depression, and Mood among Individuals with Mental Retardation. Journal of Autism and Developmental Disorders. 2003.
52. Rutter Bailey, A., & Lord, C. M. Manual of the Social Communication Questionnaire. Los Angeles, CA. 2003;
53. Constantino CPG& JN. Social Responsiveness Scale (SRS). Los Angeles, CA West Psychol Serv. 2011;(2005).
54. McIntosh DN, Miller LJ, Shyu V, Dunn W. Development and validation of the short sensory profile. Sens profile Man. 1999;61:59–73.
55. Vermulst A, Kroes G, de Meyer R, Nguyen L, Veerman JW. Opvoedingsbelastingvragenlijst (OBVL). Praktikon. 2012.
56. 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.
57. Elliott R, Wagner J, Sales CMD, Rodgers B, Alves P, Café MJ. Psychometrics of the personal questionnaire: A client-generated outcome measure. Psychol Assess. 2016.
58. Sparrow SS, Cicchetti D V., Saulnier CA. Vineland Adaptive Behavior Scales, Third Edition. San Antonio, TX: Pearson; 2016.
59. Pepperdine CR, McCrimmon AW. Test Review: Vineland Adaptive Behavior Scales, Third Edition (Vineland-3). Can J Sch Psychol. 2018;33(2).
60. Anagnostou E, Jones N, Huerta M, Halladay AK, Wang P, Scahill L, et al. Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. Autism. 2015.
61. Schmidt JD, Huete JM, Fodstad JC, Chin MD, Kurtz PF. An evaluation of the Aberrant Behavior Checklist for children under age 5. Res Dev Disabil. 2013.
62. Stoddard J, Zik J, Mazefsky CA, DeChant B, Gabriels R. The Internal Structure of the Aberrant Behavior Checklist Irritability Subscale: Implications for Studies of Irritability in Treatment-Seeking Youth With Autism Spectrum Disorders. Behav Ther. 2020;51(2).
63. Ghaleiha A, Mohammadi E, Mohammadi MR, Farokhnia M, Modabbernia A, Yekehtaz H, et al. Riluzole as an adjunctive therapy to risperidone for the treatment of irritability in children with autistic disorder: A double-blind, placebo-controlled, randomized trial. Pediatr Drugs. 2013;15(6).
64. Sansone SM, Widaman KF, Hall SS, Reiss AL, Lightbody A, Kaufmann WE, et al. Psychometric study of the aberrant behavior checklist in fragile X syndrome and implications for targeted treatment. J Autism Dev Disord. 2012.

65. Berry-Kravis E, Hessl D, Abbeduto L, Reiss AL, Beckel-Mitchener A, Urv TK. Outcome measures for clinical trials in fragile X syndrome. J Dev Behav Pediatr. 2013.
66. Budimirovic DB, Berry-Kravis E, Erickson CA, Hall SS, Hessl D, Reiss AL, et al. Updated report on tools to measure outcomes of clinical trials in fragile X syndrome. Journal of Neurodevelopmental Disorders. 2017.
67. Tate RL, Perdices M, Rosenkoetter U, Wakim D, Godbee K, Togher L, 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.
68. Hoyer D, Zorrilla EP, Cottone P, Parylak S, Morelli M, Simola N, et al. Clinical Global Impression Scales. In: Encyclopedia of Psychopharmacology. 2010.
69. Guy W. ECDEU Assessment manual for Psuchopharmacology Revised: Clinical Global Impression. Nimh. 1976;1.
70. Shapiro EG, Nestrasil I, Ahmed A, Wey A, Rudser KR, Delaney KA, et al. Quantifying behaviors of children with Sanfilippo syndrome: The Sanfilippo Behavior Rating Scale. Mol Genet Metab. 2015;114(4).
71. Sappok T, Diefenbacher A, Gaul I, Bölte S. Validity of the social communication questionnaire in adults with intellectual disabilities and suspected autism spectrum disorder. Am J Intellect Dev Disabil. 2015;120(3).
72. Brooks WT, Benson BA. The validity of the social communication questionnaire in adults with intellectual disability. Res Autism Spectr Disord. 2013;7(2).
73. Chandler S, Charman T, Baird G, Simonoff E, Loucas T, Meldrum D, et al. Validation of the Social Communication Questionnaire in a population cohort of children with autism spectrum disorders. J Am Acad Child Adolesc Psychiatry. 2007;46(10).
74. Berument SK, Rutter M, Lord C, Pickles A, Bailey A. Autism screening questionnaire: Diagnostic validity. Br J Psychiatry. 1999;175(NOV.).
75. Dunn W. Sensory profile: User's manual. San Antonio, TX: Psychological Corporation; 1999.
76. Williams ZJ, Failla MD, Gotham KO, Woynaroski TG, Cascio C. Psychometric Evaluation of the Short Sensory Profile in Youth with Autism Spectrum Disorder. J Autism Dev Disord. 2018;48(12).
77. Gaasterland CMW, Jansen-Van Der Weide MC, Weinreich SS, Van Der Lee JH. A systematic review to investigate the measurement properties of goal attainment scaling, towards use in drug trials. BMC Med Res Methodol. 2016.
78. Fisher RS. The New Classification of Seizures by the International League Against Epilepsy 2017. Vol. 17, Current Neurology and Neuroscience Reports. 2017.
79. Wilkinson M, Dumontier M, Aalbersberg I. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3 (2016). Sci data. 2016;3.
80. Araujo A, Julious S, Senn S. Understanding variation in sets of N-of-1 trials. PLoS One. 2016.
81. Neurophysiological Biomarker Toolbox [Internet]. [cited 2023 Jul 25]. Available from: http://www.nbtwiki.net/
82. The MathWorks Inc. Natick, MA, R2022a.
83. Woodcock J, LaVange LM. Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both. N Engl J Med. 2017;377(1).
84. Cummings J, Montes A, Kamboj S, Cacho JF. The role of basket trials in drug development for neurodegenerative disorders. Vol. 14, Alzheimer's Research and Therapy. 2022.

PART II

Outcome measures for rare genetic neurodevelopmental disorders and intellectual disability

Chapter 5

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

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

Submitted

Abstract

Objective: Individuals with genetic neurodevelopmental disorders (GNDs) or intellectual disability (ID) are often affected by complex neuropsychiatric comorbidities. Targeted treatments are increasingly available, but due to the heterogeneity of these patient populations, choosing a key outcome and outcome measurement instrument remains challenging. This scoping review aimed to provide an overview of outcomes and instruments used in clinical trials in GNDs and ID.

Methods: The protocol was published in the Open Science Framework. MEDLINE, PsycINFO and Cochrane CENTRAL were searched for clinical trials in individuals with GNDs and ID over the past ten years. Information was recorded on patient populations, interventions, designs, outcomes, measurement instruments, and type of reporter when applicable. Qualitative and descriptive analyses were performed.

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

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

Box 1. Definitions and abbreviations of commonly used terminology with regard to outcomes and outcome measurement instruments, adapted from the Food and Drug Administration (*).[1–4]

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

Box 1. continued

Definition	Abbreviation	Explanation
Condition-specific outcome measures		A measure capturing elements of health relevant to a particular patient group or designed for a specific patient population.
Personalized outcome measure		A measure that refers to an instrument in which the domains and/or weights are not fixed. Outcome areas are specific for each individual and the affected individual (or proxy) is involved in identifying and setting specific outcome areas. In clinical trials, these are intended for standardized evaluation of an intervention's effectiveness based on individualized problems or goals.

Introduction

Intellectual disability (ID) occurs in 1-3% of the population and is characterized by substantial limitations in both intellectual functioning and adaptive behavior, originating during the developmental period.[5–7] Exogenous factors such as an infection and birth complications may cause ID,[8] and with novel techniques such as exome and genome sequencing, a genetic etiology can be identified in up to 50% of the individuals with ID with many more awaiting diagnosis (Figure 1).[9,10] Although these genetic neurodevelopmental disorders (GNDs), including syndromic ID and neurometabolic disorders, are individually rare, collectively they are common.[11,12] In GNDs, the level of intellectual functioning is variable, ranging from normal or borderline functioning to profound ID.[13–15] Although GNDs and ID populations have often been separately studied, there is substantial overlap in patient populations.

Individuals with GND and ID are often affected by complex somatic and neuropsychiatric comorbidity, with great inter- and intra-individual variability. Neuropsychiatric manifestations typically cause the greatest burden for the affected individual, their families, and on health care systems, with a substantial clinical and economic burden.[16] The Food and Drug Administration (FDA) considers clinical outcome assessments (COAs), including patient-reported outcomes (PROs), clinician-reported outcomes (ClinROs), observer-reported outcomes (ObsROs), and performance

outcomes (PerfOs), well-defined and reliable assessments of affected individuals' symptoms, overall mental state, or how they function.[4,17,18]

Knowledge about the genetic etiology of GNDs rapidly increases and offers disorder-specific treatment options which can be targeted to the gene, protein, or downstream biological pathway.[19] It allows for personalized care, which is the implementation of etiology-drive health monitoring and treatments.[20] For neuropsychiatric manifestations, targeted treatments are underway[21–24] and guidelines are increasingly available.[25]

However, interventional research in GNDs and ID is challenging. This is due to the rarity, complexity and variability of health manifestations, even among individuals with the same disorder, as well as the heterogeneity in treatment response.[26] Other hurdles in these populations include varying cognitive and adaptive abilities, environmental factors, high rate of behavioral and emotional disturbances, a lack of stability, practice effects, and lack of consensus on the best measures within a particular construct.[27] Many outcomes have been measured in the past, but assessments of disease severity using clinical rating scales omitted patient perspectives about issues of relevance to their health. Deciding upon an appropriate outcome measure can be a daunting task, taking into account the acceptability and feasibility, and important measurement properties, such as validity, reliability, and responsiveness to change.

Noticeably, selection of outcome measures for a study has far-reaching implications. Previous trials that did not demonstrate significant clinical benefits based on the primary endpoints have been deemed 'negative' or 'failed' even though improvement on secondary endpoints or in clinical subgroups may be present,[27] as happened for clinical trials investigating the effects of Arbaclofen in Fragile X syndrome.[28,29] As such, inappropriate outcomes or outcome measurement instruments can result in negative results about the effectiveness of interventions, potentially meaning that truly effective treatments do not become available to patients and their families.[19,30]

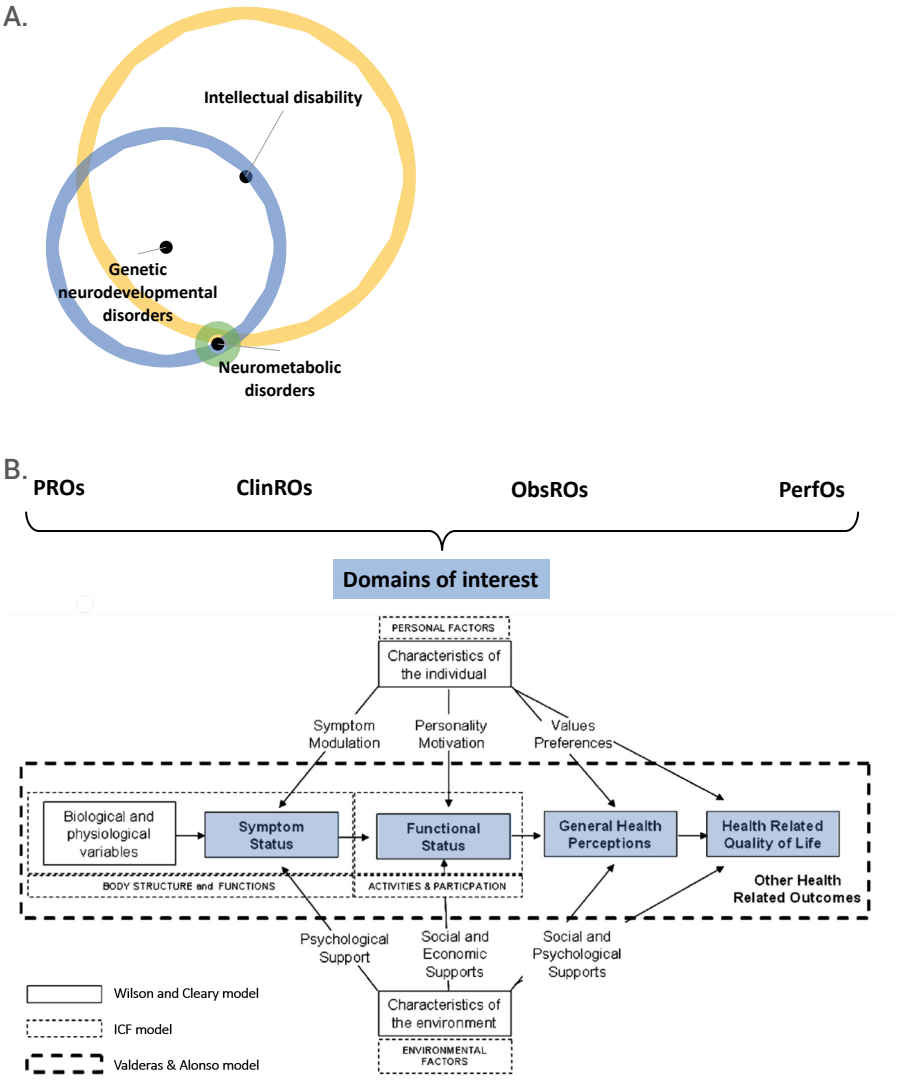


Figure 1. A) Schematic representation of target populations. Importantly, it represents an indication rather than a precise scaled proportion of target populations. **B)** Types of clinical outcome assessments (upper) and domains of interest to this review (lower; blue boxes: symptom and functional status related to neurological functioning, and the overarching concepts of perceived health and overall quality of life). This is based on the conceptual model of health outcomes from Valderas & Alonso which incorporates both the commonly used models of the ICF and Wilson & Cleary.[3,31] Figure adapted from Valderas & Alonso.

ClinRO, clinician-reported outcome; ICF, International Classification of Functioning, disability and health; ObsRO, observer-reported outcome; PerfO, performance outcome; PRO, patient-reported outcome.

The aim of this scoping review was to provide an overview of outcomes and outcome measurement instruments selected in clinical trials in individuals with ID and GNDs, measured by COAs focusing on neurological functioning, mental and social functioning, and the overarching concepts of perceived health and health-related quality of life (HR-QoL).

Methods

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

Eligibility

Clinical trials for any intervention were included in the review, including comparative studies (randomized and non-randomized), single-case trials and single-arm case series (retrospective and prospective), and trial protocols. Validation and feasibility studies, and economic evaluations were excluded. GNDs were defined as disorders with a genetic etiology affecting the nervous system in early development. GNDs associated with ID were included. ID was defined as substantial limitations in both intellectual functioning and adaptive behavior, originating during the developmental period.[5] Neurometabolic disorders, consisting of a subgroup of rare genetic hereditary conditions in which the impairment of a biochemical pathway is essential to the pathophysiology of the disease,[34] were included in case they are associated or presented with intellectual deficits/impairment. Studies were included when a participant showed intellectual impairments. Exclusion criteria included ID explicitly stated to be due to exogenous factors.

Studies were included if these used a COA, i.e., a PRO, ClinRO, ObsRO, or PerfO (Box 1; Figure 1). Condition-specific and personalized outcome measures were included as well (Box 1). Biological and physiological variables were excluded. Clinical trials with only epilepsy characteristics or

motor function as outcome without another COA about mental or social functioning, general health perceptions or HR-QoL were excluded to narrow the scope of the review. Eligible assessments included descriptions related to neurological functioning, mental and social functioning, general health perceptions or HR-QoL.

Search strategy, study selection

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

The application Rayyan was used for screening.[35] Titles and abstracts were independently screened for eligibility by six reviewers (AM, BdH, EB, LB, MB, AvE) who all screened one-sixth of the selected items, with a subsample of 10% double-screening for interrater reliability. Discrepancies were discussed until consensus was reached. Full texts were screened for eligibility, and data were independently reviewed by seven reviewers (AM, BdH, EB, LB, MB, NvS, AvE) with a sample of 10% double-reviewing for interrater reliability. Potential discrepancies were solved through discussion.

Data extraction

The following data were extracted: title, year of publication, first author, journal, countries of study, type of study, number of participants, GND/

neurometabolic disorder/heterogeneous ID of unknown cause (if neurometabolic disorder, this category was used; GNDs and neurometabolic disorders can overlap), diagnosis, patient characteristics (including age, sex, severity of ID), design, duration of trial, randomization, blinding, intervention, comparator used, type of COA, reported outcomes, outcome measurement instruments, whether it concerned a condition-specific or personalized outcome measure, type of report, number of assessments, mode of data collection, setting, and involvement of patient/parent perspectives regarding the choice of outcome measures. The reported outcomes were classified according to the most commonly used terms by the authors of the included studies. As for outcomes related to behavior, the term 'behavior' was used when general behavior was reported or when it was not further specified. Otherwise, terminology for specific behavior was reported, such as 'repetitive behavior'.

Outcome measurement instruments were classified based on the reported outcomes and information provided in the articles using a conceptual model of health outcomes from Valderas & Alonso, which is a combination of the classification system of Wilson and Cleary and the International Classification of Functioning, Disability and Health (Figure 1B).[3,36] Domains included symptoms, physical function, mental function, social function, general health perceptions, and HR-QoL. Additionally, cognitive function was included as a separate domain to better distinguish mental health domains, considering the target population.

Results

Of 4,507 identified citations, 312 studies met the inclusion criteria, with 251 research articles and 61 trial protocols in registers.

Study characteristics, population, interventions, methodology

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

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

Diagnosis	N
Intellectual disability of unstated etiology [a]	143
Down syndrome	33
Fragile X syndrome	23
Prader-Willi syndrome	23
Tuberous Sclerosis Complex	17
Mucopolysaccharidosis [b,c]	12
Neurofibromatosis type 1	9
Rett syndrome	8
Phenylketonuria [c]	7
Angelman syndrome	6
22q11.2 deletion syndrome	5
Niemann-Pick disease type C	4
Fragile X premutation-associated conditions	3
Smith-Magenis syndrome	3
1p36 deletion syndrome	2
Coffin-Siris syndrome	2
Cornelia de Lange syndrome	2
Kabuki syndrome	2
Metachromatic leukodystrophy	2
Phelan-McDermid syndrome	2
Succinic semialdehyde dehydrogenase deficiency	2
Williams syndrome	2
XYY syndrome [c]	2
Aicardi-Goutières Syndrome	1
Alpha-mannosidosis	1
Classic galactosemia	1
Congenital lipomatous overgrowth, vascular malformations, and epidermal nevi (CLOVE syndrome)	1
Cyclin-dependent kinase like-5 deficiency disorder	1
Klippel-Trenaunay syndrome	1
Leigh syndrome	1
Mitochondrial disorders	1
Pantothemate kinase-associated neurodegeneration	1
Phosphomannomutase-2 congenital disorder of glycosylation (PMM2-CDG)	1
Propionic acidemia [c]	1
Smith-Lemli-Opitz syndrome	1

a One study included a participant with mucopolysaccharidosis type IV (Morquio A syndrome) which was however not considered associated with the diagnosed syndrome.

b Mucopolysaccharidosis type I (n=3), type II (n=1), type IIIA (n=4), type IIIB (n=3).

c Disorders that are not always associated with intellectual disability (i.e., due to advanced screening and therapies), but (some) participants included in these studies were affected with intellectual impairments.

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

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

Reported outcomes

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

Table 2. Reported outcomes and number of outcome measurement instruments used, clustered according to frequency of use.

Reported outcomes	Frequency	Number of different outcome measurement instruments used	PRO	ClinRO	ObsRO	PerfO
Cognitive function	333	141	33	16	15	269
(HR-)QoL	74	23	72	2	0	0
Aberrant behavior (e.g., challenging / maladaptive / dysfunctional / destructive behavior / (severe) behavioral problems / manifestations)	64	18	59	2	3	0
(Clinical) global impression (including severity/ improvement)	64	16	15	49	0	0
Communication	59	35	11	0	13	35
Behavior (general / not specified)	45	23	36	3	6	0
Adaptive behavior	33	5	29	2	2	0
Depression and mood disorders	33	15	27	3	3	0
Autism	31	14	15	11	5	0

Table 2. continued

Reported outcomes	Frequency	Number of different outcome measurement instruments used	PRO	ClinRO	ObsRO	PerfO
Anxiety	27	15	26	1	0	0
Mental health (e.g., global, well-being, feelings, psychological wellness/distress, symptoms of mental disorder)	27	16	18	7	2	0
Social behavior	27	13	22	1	4	0
Sleep	26	12	19	0	7	0
Other / unclear	22	17	14	5	3	0
Motor function	20	19	3	7	2	8
Participation	18	11	13	0	5	0
Personalized goals	14	6	13	1	0	0
Emotion regulation	13	11	12	0	1	0
Activity	11	7	9	0	2	0
Syndrome-specific symptoms	10	8	5	5	0	0
Academic skills	9	9	2	0	1	6
Attention	9	9	5	1	3	0
Repetitive behavior	9	2	8	1	0	0
Aggression	9	1	1	0	5	0
Anger	8	4	8	0	0	0
Hyperphagia	8	1	7	1	0	0
Pain	7	6	7	0	0	0
Irritability	7	1	7	0	0	0
Self-injurious behavior	6	5	5	1	0	0
Eating behavior	5	4	5	0	0	0
Epilepsy	5	2	5	0	0	0
Psychiatric symptoms	5	3	1	4	0	0
Self-efficacy	5	5	3	1	1	0
Self-esteem	5	1	5	0	0	0
Ataxia	4	3	0	4	0	0
Neurological function	4	1	0	4	0	0
Social support	4	3	4	0	0	0
Stress	4	4	4	0	0	0
Substance use	4	3	4	0	0	0

Table 2. continued

Reported outcomes	Frequency	Number of different outcome measurement instruments used	PRO	ClinRO	ObsRO	PerfO
Concerns	3	3	1	2	0	0
Psychosocial function	3	2	3	0	0	0
Post-traumatic stress disorder	3	2	3	0	0	0
Alertness	2	1	2	0	0	0
Coping behavior	2	2	2	0	0	0
Mentalizing abilities	2	2	0	0	0	2
Obsession and compulsivity	2	1	0	2	0	0
Resilience	2	2	2	0	0	0
Suicide	2	2	2	0	0	0
(Self-)compassion	1	1	1	0	0	0
Acceptance	1	1	1	0	0	0
Apathy	1	1	1	0	0	0
Confusion	1	1	1	0	0	0
Dysarthria	1	1	0	1	0	0
Dyskinesia	1	1	0	1	0	0
Dystonia	1	1	0	0	1	0
Empowerment	1	1	1	0	0	0
Hyperactivity	1	1	1	0	0	0
Life events	1	1	1	0	0	0
Psychosis	1	1	1	0	0	0
Satisfaction	1	1	1	0	0	0
Self-determination	1	1	1	0	0	0
Combined terms*						
Attention; Hyperactivity; Impulsivity	11	6	10	1	0	0
Anxiety; Depression and mood disorders	7	2	7	0	0	0
Cognitive function; Adaptive behavior	5	4	2	1	0	2
Cognitive function; Motor function	5	4	1	0	2	2
Behavior; Emotion regulation	4	2	4	0	0	0
Self-efficacy; Social support	4	2	4	0	0	0
Cognitive function; Motor function; Communication	3	3	0	1	1	1
Behavior; Cognitive function	2	1	0	2	0	0

Table 2. continued

Reported outcomes	Frequency	Number of different outcome measurement instruments used	PRO	ClinRO	ObsRO	PerfO
Cognitive function; Communication	2	2	0	0	0	2
Communication; Activity; Social behavior	2	1	2	0	0	0
Depression and mood disorders; behavior	2	1	2	0	0	0
Emotion regulation; Social behavior	2	1	2	0	0	0
Emotion regulation; Social behavior; Eating behavior	2	1	2	0	0	0
Aggression; Social behavior	1	1	1	0	0	0
Anxiety; Irritability	1	1	1	0	0	0
Cognitive function; Behavior	1	1	0	0	0	1
Cognitive function; Communication; Social behavior	1	1	1	0	0	0
Cognitive function; Emotion regulation	1	1	0	0	0	1
Cognitive function; Global impression	1	1	0	1	0	0
Cognitive function; Motor function; Emotion regulation	1	1	0	0	0	1
Communication; Social behavior	1	1	1	0	0	0
Eating behavior; Mental health	1	1	1	0	0	0
Emotion regulation; Social behavior; Activity	1	1	1	0	0	0
Irritability; Hyperactivity	1	1	1	0	0	0
Mental health; Autism	1	1	0	1	0	0
Pain; Mental health; Social behavior	1	1	0	0	1	0
Pain; Stress; Social behavior	1	1	0	1	0	0
Satisfaction; Mental health	1	1	1	0	0	0

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

Outcome measurement instruments

Of the 457 different outcome measurement instruments that were identified, 213 (47%) were classified as instruments for PROs, 54 (12%) as ClinROs, 48 (11%) as ObsROs, and 157 (34%) as PerfOs. There were 288 (63%) outcome measurement instruments that were used in only one clinical trial. Another 16 (4%) outcome measurement instruments were self-designed for the particular trial, classified as instrument for PROs (n=12) and ObsROs (n=4). Instruments for PerfOs measured cognitive function and physical function.

There were 18 condition-specific outcome measurement instruments used in 30 (10%) clinical trials in total, including instruments for Down syndrome, Prader-Willi syndrome, phenylketonuria, mitochondrial disease, Rett syndrome, Fragile X syndrome, Niemann-Pick disease type C, and phosphomannomutase deficiency congenital disorder of glycosylation. Two condition-specific outcome measurement instruments were designed ad-hoc for the specific trial.

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

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

Discussion

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

to enable evidence-based general ID care and measuring effectiveness of innovative therapies.

Reported outcomes

From a total of 312 studies, there were 438 different reported outcomes clustered into 91 different outcomes. We encountered large differences in terminology for similar constructs, such as 'aberrant behaviors', 'challenging behaviors', 'behavioral problems', and 'severe behavioral manifestations'. This may conflict with generalizability and clarity among clinical trials, demonstrating the need for semantic harmonization. Similarly, overlap in PROs across the International Consortium for Health Outcomes Measurement (ICHOM) standard sets was recently examined, identifying 307 different PROs referring to 22 unique PRO concepts.[37] Furthermore, (HR-)QoL was reported in 74 clinical trials, using 23 different instruments. Although HR-QoL is an important outcome, this broad, abstract, and multidimensional concept can cover different concepts, obscuring the construct to be measured. According to the FDA, an HR-QoL measure should at a minimum capture physical, psychological (including emotional and cognitive), and social functioning.[1]

Outcome measurement instruments

We identified 457 different outcome measurement instruments to measure patient-reported outcomes (n=213), clinician-reported outcomes (n=54), observer-reported outcomes (n=48), and performance outcomes (n=157), with 288 instruments (63%) only used in one clinical trial in the past decade. The large number of different instruments used in clinical trials is not surprising, considering the heterogeneity in levels of intellectual functioning, patients and researcher preferences, availability of instruments that are appropriate to specific conditions, and regional preferences. Furthermore, for novel drugs with yet unknown efficacy, multiple domains might be studied requiring different instruments, to investigate effectiveness and identify potential subgroups who benefit most from the intervention. This is also reflected by the large amount of ad-hoc designed symptom- and condition-specific instruments, hampering extrapolation and interpretation of the results. Yet, it is laborious to examine validity, reliability, and responsiveness of so many instruments. It underlines the need for consensus on outcomes and instruments, such as the Outcome Measures Working Groups and

expert groups convened by the NIH,[38,39] the ERICA PROMs Repository (Endo-ERN), and the establishment and validation of the National Institutes of Health Toolbox Cognitive Battery (NIH-TCB) for individuals with ID.[27,41]

Type of reporter

Although instruments are generally developed as one specific type of COA, similar instruments were completed by different types of reporters (e.g., a ClinRO instrument used as ObsRO by parents). Furthermore, proxy-reports were substantially more used (39%) than self-(assisted) reports (23%). Although the use of proxy-reports is not surprising in populations with ID, the validity of proxy reflections of unobservable internal states (e.g., anxiety or depression) is limited, as the personal perspective can only truly be understood by the individual's self-report.[42] Proxy-raters often assess (HR-) QoL worse compared to individuals themselves, indicating bias.[43–49] PROs may thus be difficult to measure by proxy-reports,[50] although still providing valuable information.[51] It has been suggested that adolescents with ID can reliably report on their mental health status, with instruments appropriate to their age, cognitive and visuospatial functioning.[52,53] However, a recent study illustrated that there is currently no self-report instrument available that is recommended for assessing (HR-)QoL and subjective wellbeing of adolescents with ID, based on psychometric evidence.[42] Novel methods are upcoming such as experience sampling methods with the use of apps and ID-friendly instruments.[54]

The emergence of condition-specific and personalized outcome measurement instruments

In an attempt to target more specific phenotypes without the need for multiple tools to measure the impact of the disease,[55,56] condition-specific (n=18) and personalized (n=6) outcome measurement instruments, and tools particularly designed for a specific trial (n=16) were used. Condition-specific PROMs have also been developed due to unavailability of proxy-versions for adults with ID and criticism on appropriateness of the existing instrument's content and measurement properties for the target population.[55,57] Such instruments might contain more relevant items to complete, increasing acceptability among affected individuals. However, results might be difficult to generalize or interpret. Furthermore, it is not

feasible and desirable to use condition-specific PROMs for more than 7000 rare disorders.[58] It may also not be necessary, as research has shown that PRO domains that patients consider important are very similar among patient populations.[37]

Generic instruments have the advantage of allowing comparison of outcomes between different disease (sub)groups. Generally, all individuals want to feel and function well, such as living without symptoms and being able to carry out daily activities. Feelings and functions can be affected by different health conditions, and these can result to similar problems with considerable overlap in relevant PROs across conditions, which could be measured with one set of generic outcome measures across conditions.[36,37] Methodological innovations, such as item response theory (IRT), have been used to develop PROMs with good measurement properties that are applicable across different health conditions, such as Patient-Reported Outcomes Measurement Information System (PROMIS®).[59–61] IRT-based item banks are large sets of calibrated questions measuring the same construct, enabling efficient measurement through short forms or computerized adaptive testing (CAT).[41,62] This provides a valuable solution, since redundant items for specific individuals will be minimized, increasing relevance and efficiency.[63]

To ensure relevance, personalized outcome measurement instruments have gained emerging interest, especially for rare and heterogeneous patient populations since health manifestations are often specific, variable and complex.[64] Instruments such as Goal Attainment Scaling enable focusing on personal goals and abilities.[65] Additionally, by including outcomes that are specifically relevant to the affected individual, treatment adherence might be enhanced as well.[66] Also regulatory agencies have increasing interest in the relevance of what is being measured,[67] as treatment effects might be statistically significant, but not clinically or socially relevant, or vice versa.[68]

Recommendations for selecting outcomes and instruments

In order to measure what matters to patients, several important factors should be taken into account when selecting outcome measurement instruments in clinical trials (Table 3).[69] First, relevance to the patient should be ensured, which contributes to recruitment as well as treatment

compliance.[27,66] Affected individuals and representatives of the target population should be formally involved in the choice of measured outcomes, while now involvement was mentioned in only 2% of the clinical trials.

When selecting instruments, their acceptability, feasibility, and measurement properties should be taken into account (Table 3). For already over-burdened caregivers, outcome measurement instruments can be time-consuming to complete, and are often experienced as confronting due to inappropriateness of questions, leading to poor acceptability.[55] Furthermore, it is recommended to attempt to include (user-friendly) PROMs to acquire information from the patient perspective, as also encouraged by regulatory authorities such as the FDA and EMA.[70,71]

Table 3. Recommendations, as provided by the authors, with regard to selecting outcomes and outcome measurement instruments in clinical trials for individuals with genetic neurodevelopmental disorders and/or intellectual disability.

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

ERICA, European Rare Disease Research Coordination and Support Action consortium; NIH-TCB, National Institutes of Health Toolbox Cognitive Battery (NIH-TCB); PROM, patient-reported outcome measure; PROMIS, Patient-Reported Outcomes Measurement Information System.

Future outcome measure landscape

Because of the overgrowth of available outcome measurement instruments, clinical researchers need guidance in choosing appropriate outcome measures in clinical trials. Regulatory agencies, such as the European Medicines Agency (EMA) and FDA, encourage maintaining consistency in assessment methods and are placing focus on capturing the patient experience, but poorly defined PRO objectives have hindered the utility of PROs in regulatory decisions.[67,72] A core outcome set or generic measure with disorder- or comorbidity-specific extensions may provide a solution to ensure generalizability and interpretation, and effectively target specific phenotypes in individuals with GNDs and ID. To move from this 'mess of measures' to a more unified approach for future interventional research for GNDs and ID, the field could take the following steps:

- Reach (international) consensus on outcomes (e.g., Delphi procedure) and establish a core outcome set for individuals with GNDs and ID: terminology and constructs should be relevant, clear, harmonized and operationalized, in collaboration with affected individuals, caregivers, and (methodological and clinical) experts.
- Reach (international) consensus on the most suitable instruments to be selected per outcome, taking into account relevance, applicability, patient preferences, validity, reliability, responsiveness to change, and language and culture barriers. Some instruments may need to be adapted to individuals with ID.
- Implement the core outcome set or (ID-friendly) generic measure(s) with appropriate versions for different levels of ID. This could be extended with disorder- or comorbidity-specific measures (e.g., symptom checklists) to cover relevant condition-specific aspects.

Strengths and limitations

This scoping review is the first rigorous overview of outcomes and outcome measurement instruments used in clinical trials in GNDs and ID, examining the broad array of outcomes related to health manifestations common in

these patient populations, using state of the art classifications. However, when conducting this review, we faced some challenges. We initially aimed to cluster the outcomes and outcome measurement instruments according to the Valderas & Alonso model.[3] Domain assessment has rather been an indication, as instruments should ideally be assessed per subscale (unidimensional), which was not feasible due to the enormous amount of different outcome measurement instruments. Furthermore, the terminology used for the outcomes and outcome measurement instruments was often unclear, lacking, or inconsistently reported. We clustered reported outcomes based on frequency of used terminology, and thus do not refer to a standardized terminology. Finally, we cannot recommend specific outcome measurement instruments, since psychometric properties were not investigated in this review.

Conclusion

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

Funding

This work was supported financially by Stichting Wetenschappelijk Onderzoek 's Heeren Loo. The funder had no role in the design and conduct of the study, preparation of the review, or approval of the manuscript.

Data Availability

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

Competing interests

The authors declare that they have no competing interest.

Authors' contributions

AM and AvE were involved in the study concept and design, had a major role in the acquisition of data, analysis and interpretation of data, and drafted the manuscript. NvS had a major role in the acquisition, analysis and interpretation of data, and revised the manuscript. BdH, MB, and EB were involved in the study concept and design, had a major role in the acquisition of data, and revised the manuscript. DK had a major role in processing and the analysis of data. LB had a major role in the acquisition of data, and revised the manuscript. LH was involved in the study concept and design and interpretation of data. CT was involved in the study design and concept and interpretation of data, revised the manuscript for content. CS and FW were involved in the study concept and revised the manuscript. JD performed the search strategy and contributed to the analysis. DH revised the manuscript for content.

References

1. Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims: Draft guidance. Health Qual Life Outcomes. 2006;4.
2. Churruca K, Pomare C, Ellis LA, Long JC, Henderson SB, Murphy LED, et al. Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues. Vol. 24, Health Expectations. 2021.
3. Valderas JM, Alonso J. Patient reported outcome measures: A model-based classification system for research and clinical practice. Qual Life Res. 2008;17(9).
4. FDA Clinical Outcome Assessments [Internet]. [cited 2023 Jul 11]. Available from: https://www.fda.gov/about-fda/cdrh-patient-science-and-engagement-program/clinical-outcome-assessments-coas-medical-device-decision-making.
5. American Psychiatric Association. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders Fifth Edition. Arlington. 2013.
6. Leonard H, Wen X. The epidemiology of mental retardation: Challenges and opportunities in the new millennium. Vol. 8, Mental Retardation and Developmental Disabilities Research Reviews. 2002.
7. Pinchefsky E, Shevell M. Intellectual Disabilities and Global Developmental Delay. In: Handbook of DSM-5 Disorders in Children and Adolescents. 2017.
8. Kvarnung M, Nordgren A. Intellectual disability & rare disorders: A diagnostic challenge. In: Advances in Experimental Medicine and Biology. 2017.
9. Wang J, Wang Y, Wang L, Chen WY, Sheng M. The diagnostic yield of intellectual disability: Combined whole genome low-coverage sequencing and medical exome sequencing. BMC Med Genomics. 2020;13(1).
10. Pekeles H, Accogli A, Boudrahem-Addour N, Russell L, Parente F, Srour M. Diagnostic Yield of Intellectual Disability Gene Panels. Pediatr Neurol. 2019;92.
11. Castrén E, Elgersma Y, Maffei L, Hagerman R. Treatment of Neurodevelopmental disorders in Adulthood. J Neurosci. 2012;32(41).
12. Levy G, Barak B. Postnatal therapeutic approaches in genetic neurodevelopmental disorders. 3:2021. Available from: https://doi.org/10.4103/1673-5374.293133.
13. Ferrari M. Borderline intellectual functioning and the intellectual disability construct. Vol. 47, Intellectual and Developmental Disabilities. 2009.
14. Swillen A, McDonald-Mcginn D. Developmental trajectories in 22q11.2 deletion syndrome. Am J Med Genet Part C Semin Med Genet. 2015;169(2).
15. Green Snyder LA, D'Angelo D, Chen Q, Bernier R, Goin-Kochel RP, Wallace AS, et al. Autism Spectrum Disorder, Developmental and Psychiatric Features in 16p11.2 Duplication. J Autism Dev Disord. 2016;46(8).
16. Cannizzo S, Lorenzoni V, Palla I, Pirri S, Trieste L, Triulzi I, et al. Rare diseases under different levels of economic analysis: Current activities, challenges and perspectives. Vol. 4, RMD Open. 2018.
17. Walton MK, Powers JH, Hobart J, Patrick D, Marquis P, Vamvakas S, et al. Clinical Outcome Assessments: Conceptual Foundation-Report of the ISPOR Clinical Outcomes Assessment-Emerging Good Practices for Outcomes Research Task Force. Value Heal. 2015;18(6).
18. Gnanasakthy A, Qin S, Norcross L. FDA Guidance on Selecting, Developing, or Modifying Fit-for-Purpose Clinical Outcome Assessments: Old Wine in a New Bottle? Vol. 16, Patient. 2023.
19. Overwater IE, Rietman AB, van Eeghen AM, de Wit MCY. Everolimus for the treatment of refractory seizures associated with tuberous sclerosis complex (TSC): Current perspectives. Therapeutics and Clinical Risk Management. 2019.
20. Van Eeghen AM, Bruining H, Wolf NI, Bergen AA, Houtkooper RH, Van Haelst MM, et al. Personalized medicine for rare neurogenetic disorders: can we make it happen? Cold Spring Harb Mol Case Stud. 2022;8(2):a006200.

21. Jacquemont S, Berry-Kravis E, Hagerman R, Von Raison F, Gasparini F, Apostol G, et al. The challenges of clinical trials in fragile X syndrome. Psychopharmacology. 2014.
22. Jansen-Van Der Weide MC, Gaasterland CMW, Roes KCB, Pontes C, Vives R, Sancho A, et al. Rare disease registries: Potential applications towards impact on development of new drug treatments. Orphanet J Rare Dis. 2018.
23. Heussler H, Cohen J, Silove N, Tich N, Bonn-Miller MO, Du W, et al. A phase 1/2, open-label assessment of the safety, tolerability, and efficacy of transdermal cannabidiol (ZYN002) for the treatment of pediatric fragile X syndrome. J Neurodev Disord. 2019.
24. Elgersma Y, Sonzogni M. UBE3A reinstatement as a disease-modifying therapy for Angelman syndrome. Vol. 63, Developmental Medicine and Child Neurology. 2021.
25. Northrup H, Aronow ME, Bebin EM, Bissler J, Darling TN, de Vries PJ, et al. Updated International Tuberous Sclerosis Complex Diagnostic Criteria and Surveillance and Management Recommendations. Pediatr Neurol. 2021;123.
26. Müller AR, Brands MM, van de Ven PM, Roes KC, Cornel MC, van Karnebeek CD, et al. The Power of 1: Systematic Review of N-of-1 Studies in Rare Genetic Neurodevelopmental Disorders. Neurology. 2021.
27. Hessl D, Sansone SM, Berry-Kravis E, Riley K, Widaman KF, Abbeduto L, et al. The NIH Toolbox Cognitive Battery for intellectual disabilities: Three preliminary studies and future directions. J Neurodev Disord. 2016;8(1).
28. Berry-Kravis EM, Hessl D, Rathmell B, Zarevics P, Cherubini M, Walton-Bowen K, et al. Effects of STX209 (Arbaclofen) on neurobehavioral function in children and adults with fragile X syndrome: A randomized, controlled, phase 2 trial. Sci Transl Med. 2012;4(152).
29. Berry-Kravis E, Hagerman R, Visootsak J, Budimirovic D, Kaufmann WE, Cherubini M, et al. Arbaclofen in fragile X syndrome: Results of phase 3 trials. J Neurodev Disord. 2017;9(1).
30. Erickson CA, Davenport MH, Schaefer TL, Wink LK, Pedapati E V., Sweeney JA, et al. Fragile X targeted pharmacotherapy: Lessons learned and future directions. Vol. 9, Journal of Neurodevelopmental Disorders. 2017.
31. Wilson IB. Linking Clinical Variables With Health-Related Quality of Life. JAMA. 1995;273(1).
32. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation. Vol. 169, Annals of Internal Medicine. 2018.
33. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Rev Esp Nutr Humana y Diet. 2016.
34. Poletti V, Biffi A. Gene-Based Approaches to Inherited Neurometabolic Diseases. Vol. 30, Human Gene Therapy. 2019.
35. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016.
36. Dodd S, Clarke M, Becker L, Mavergames C, Fish R, Williamson PR. A taxonomy has been developed for outcomes in medical research to help improve knowledge discovery. J Clin Epidemiol. 2018;96.
37. Terwee CB, Zuidgeest M, Vonkeman HE, Cella D, Haverman L, Roorda LD. Common patient-reported outcomes across ICHOM Standard Sets: the potential contribution of PROMIS®. BMC Med Inform Decis Mak. 2021;21(1).
38. Budimirovic DB, Berry-Kravis E, Erickson CA, Hall SS, Hessl D, Reiss AL, et al. Updated report on tools to measure outcomes of clinical trials in fragile X syndrome. Journal of Neurodevelopmental Disorders. 2017.
39. Esbensen AJ, Hooper SR, Fidler D, Hartley SL, Edgin J, D'Ardhuy XL, et al. Outcome measures for clinical trials in down syndrome. Am J Intellect Dev Disabil. 2017;122(3).
40. Endo-ERN. The ERICA Patient Reported Outcome Measures (PROMs) Repository [Internet]. [cited 2023 Jun 3]. Available from: https://endo-ern.eu/the-erica-patient-reported-outcome-measures-proms-repository/
41. Shields RH, Kaat AJ, McKenzie FJ, Drayton A, Sansone SM, Coleman J, et al. Validation of the NIH Toolbox Cognitive Battery in intellectual disability. Neurology. 2020.

42. Maguire S, Davison J, McLaughlin M, Simms V, Bunting B. Exploring the psychometric properties of self-report instruments used to measure health-related quality of life and subjective wellbeing of adolescents with intellectual disabilities: A Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) systematic review. J Appl Res Intellect Disabil [Internet]. 2023 Apr 26; Available from: http://www.ncbi.nlm.nih.gov/pubmed/37101341.
43. Burks HB, Des Bordes JKA, Chadha R, Holmes HM, Rianon NJ. Quality of life assessment in older adults with dementia: A systematic review. Vol. 50, Dementia and Geriatric Cognitive Disorders. 2021.
44. Griffiths AW, Smith SJ, Martin A, Meads D, Kelley R, Surr CA. Exploring self-report and proxy-report quality-of-life measures for people living with dementia in care homes. Qual Life Res. 2020;29(2).
45. Alcantara J, Ohm J, Alcantara J. Comparison of pediatric self reports and parent proxy reports utilizing PROMIS: Results from a chiropractic practice-based research network. Complement Ther Clin Pract. 2017;29.
46. Janse AJ, Gemke RJBJ, Uiterwaal CSPM, Van Der Tweel I, Kimpen JLL, Sinnema G. Quality of life: Patients and doctors don't always agree: A meta-analysis. J Clin Epidemiol. 2004;57(7).
47. Lunsky Y, Bramston P. A preliminary study of perceived stress in adults with intellectual disabilities according to self-report and informant ratings. J Intellect Dev Disabil. 2006;31(1).
48. Morrow AM, Hayen A, Quine S, Scheinberg A, Craig JC. A comparison of doctors', parents' and children's reports of health states and health-related quality of life in children with chronic conditions. Child Care Health Dev. 2012;38(2).
49. Janssen CGC, Schuengel C, Stolk J. Perspectives on quality of life of people with intellectual disabilities: The interpretation of discrepancies between clients and caregivers. Vol. 14, Quality of Life Research. 2005.
50. Ediebah DE, Reijneveld JC, Taphoorn MJB, Coens C, Zikos E, Aaronson NK, et al. Impact of neurocognitive deficits on patient–proxy agreement regarding health-related quality of life in low-grade glioma patients. Qual Life Res. 2017;26(4).
51. Upton P, Lawford J, Eiser C. Parent–child agreement across child health-related quality of life instruments: a review of the literature. Qual Life Res. 2008;17:895–913.
52. Davison J, Maguire S, McLaughlin M, Simms V. Involving adolescents with intellectual disability in the adaptation of self-reported subjective well-being measures: participatory research and methodological considerations. J Intellect Disabil Res. 2022;66(7).
53. Ingerski LM, Modi AC, Hood KK, Pai AL, Zeller M, Piazza-Waggoner C, et al. Health-Related Quality of Life Across Pediatric Chronic Conditions. J Pediatr. 2010;156(4).
54. Kooijmans R, Mercera G, Langdon PE, Moonen X. The Adaptation of Self-Report Measures to the Needs of People With Intellectual Disabilities: A Systematic Review. Clin Psychol Sci Pract. 2022;29(3).
55. Kang S, Jones A, Shaffer RC, Erickson CA, Schmitt LM. Developing Improved Outcome Measures in FXS: Key Stakeholder Feedback. SSRN Electron J. 2023.
56. Adang LA, Gavazzi F, Jawad AF, Cusack S V., Kopin K, Peer K, et al. Development of a neurologic severity scale for Aicardi Goutières Syndrome. Mol Genet Metab. 2020;130(2).
57. Müller AR, Luijten MAJ, Haverman L, De Ranitz-Greven WL, Janssens P, Rietman AB, et al. Understanding the impact of tuberous sclerosis complex: development and validation of the TSC-PROM. BMC Med. 2023;21(298).
58. Rode J. Rare diseases: understanding this public health priority. Paris: EURORDIS. 2005.
59. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. Med Care. 2007;45(5 SUPPL. 1).
60. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol. 2010;63(11).
61. Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. Vol. 7, Patient. 2014.

62. Luijten MAJ, van Litsenburg RRL, Terwee CB, Grootenhuis MA, Haverman L. Psychometric properties of the Patient-Reported Outcomes Measurement Information System (PROMIS®) pediatric item bank peer relationships in the Dutch general population. Qual Life Res. 2021;30(7).
63. Fries JF, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and promis: Assessment of physical function. J Rheumatol. 2014;41(1).
64. 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.
65. Müller AR, Zinkstok JR, Rommelse NNJ, van de Ven PM, Roes KCB, Wijburg FA, et al. Methylphenidate for attention-deficit/hyperactivity disorder in patients with Smith–Magenis syndrome: protocol for a series of N-of-1 trials. Orphanet J Rare Dis. 2021;16(1).
66. Gaasterland CMW, Jansen-van der Weide MC, Vroom E, Leeson-Beevers K, Kaatee M, Kaczmarek R, et al. The POWER-tool: Recommendations for involving patient representatives in choosing relevant outcome measures during rare disease clinical trial design. Health Policy (New York). 2018.
67. Crossnohere NL, Brundage M, Calvert MJ, King M, Reeve BB, Thorner E, et al. International guidance on the selection of patient-reported outcome measures in clinical trials: a review. Vol. 30, Quality of Life Research. 2021.
68. Pogrow S. How Effect Size (Practical Significance) Misleads Clinical Practice: The Case for Switching to Practical Benefit to Assess Applied Research Findings. Am Stat. 2019;73(sup1).
69. Morel T, Cano SJ. Measuring what matters to rare disease patients - Reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. Orphanet J Rare Dis. 2017;12(1).
70. Matts ST, Webber CM, Bocell FD, Caldwell B, Chen AL, Tarver ME. Inclusion of patient-reported outcome instruments in US FDA medical device marketing authorizations. J Patient-Reported Outcomes. 2022;6(1).
71. US Food and Drug Administration. Guidance for Industry Patient-Reported Outcome Measures Use in Medical Product Development to Support Labeling Claims. Clinical/Medical Federal Register. 2009.
72. Calvert M, King M, Mercieca-Bebber R, Aiyegbusi O, Kyte D, Slade A, et al. SPIRIT-PRO Extension explanation and elaboration: guidelines for inclusion of patient-reported outcomes in protocols of clinical trials. BMJ Open. 2021;11(6).

Supplementary figures

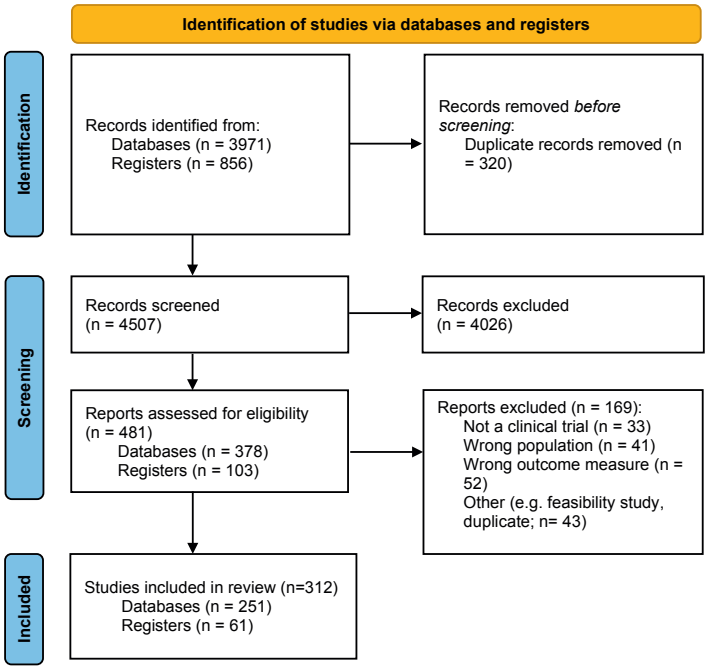


Figure 1. PRISMA Flowchart.
PRISMA = Preferred Reporting Items for Systematic Reviews and Meta-Analysis.

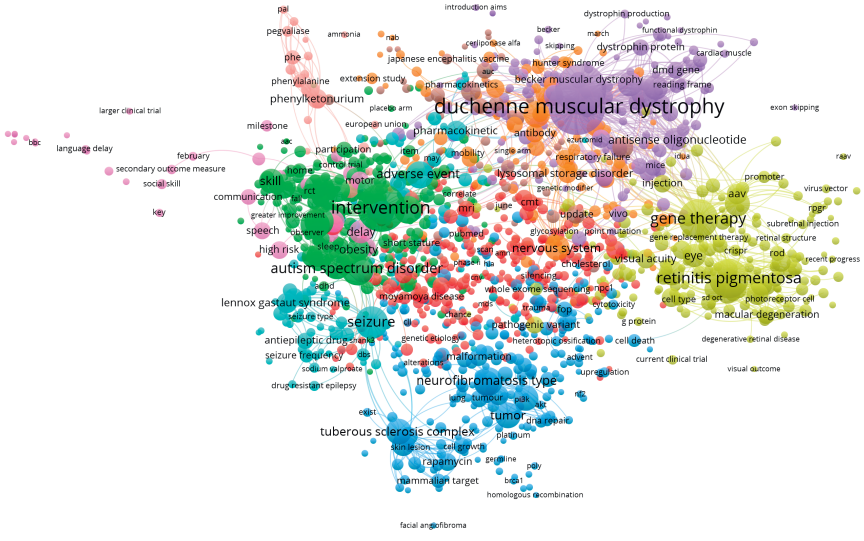


Figure 2. Results of VOSviewer to visually identify potentially irrelevant terms eligible for exclusion in order to enhance precision of search results.

Chapter 6

Understanding the impact of tuberous sclerosis complex: development and validation of the TSC-PROM

Annelieke R. Müller, Michiel A.J. Luijten, Lotte Haverman, Wendela L. de Ranitz-Greven, Peter Janssens, André B. Rietman, Leontine W. ten Hoopen, Laura C.G. de Graaff, Marie-Claire Y. de Wit, Anna C. Jansen, Tanjala Gipson, Jamie K. Capal, Petrus J. de Vries, Agnies M. van Eeghen

BMC Medicine, 2023, *21:298.*

Abstract

Background: Tuberous Sclerosis Complex (TSC) is a rare and complex genetic disorder, associated with tumor growth in various organ systems, epilepsy, and a range of neuropsychiatric manifestations including intellectual disability. With improving patient-centered care and targeted therapies, patient-reported outcome measures (PROMs) are needed to measure the impact of TSC manifestations on daily functioning. The aim of this study was to develop a TSC-specific PROM for adults that captures the impact of TSC on physical functions, mental functions, activity and participation, and the social support individuals with TSC receive, called the TSC-PROM.

Methods: COSMIN methodology was used to develop a self-reported and proxy-reported version. Development and validation consisted of the following studies: PROM development, content validity, structural validity, internal consistency, and construct validity. The International Classification of Functioning and Disability was used as a framework. Content validity was examined by a multidisciplinary expert group and cognitive interview study. Structural and construct validity, and internal consistency were examined in a large cohort, using confirmatory factor analysis, hypotheses testing, and Cronbach's alpha.

Results: The study resulted in an 82-item self version and 75-item proxy version of the TSC-PROM with four subscales (physical functions 18 and 19 items, mental functions 37 and 28 items, activities and participation 13 and 14 items, social support 13 items, for self version and proxy version respectively). Sufficient results were found for structural validity with sufficient unidimensionality for each subscale. With regard to construct validity, 82% of the hypotheses were met for the self version and 59% for the proxy version. The PROM showed good internal consistency (Cronbach's alpha 0.78-0.97).

Conclusions: We developed a PROM for adults with TSC, named TSC-PROM, showing sufficient evidence for reliability and validity that can be used in clinical and research settings to systematically gain insight into their experiences. It is the first PROM in TSC that addresses the impact of specific TSC manifestations on functioning, providing a valuable, patient-centered addition to the current clinical outcomes.

Background

Tuberous Sclerosis Complex (TSC) is a rare autosomal dominant genetic disorder with a prevalence of 1 in 6,000, caused by pathogenic variants in the *TSC1* or *TSC2* genes.[1] TSC is characterized by benign tumor growth in various organ systems, including the skin, kidneys, lungs, heart, and brain.[2] Epilepsy is a common feature of TSC and is often present in the first year of life (80%).[3] In addition, TSC is associated with varying degrees of intellectual disabilities (ID) (50%)[4] and TSC-associated neuropsychiatric disorders (TAND) (90%),[5] which encompass psychiatric, behavioral, intellectual, neuropsychological, academic and psychosocial manifestations.[3,4] The severity of TSC manifestations can vary greatly but health perception and functioning are often severely impaired.[6–9]

With improved healthcare, the largest population with TSC is now adult. Thus far, little is known of the burden and restrictions experienced by adults with TSC and the impact of TSC on functioning. As there is great variability in the severity of organ-specific involvement per life phase,[2] adult care is often variable and fragmented, including gaps in care for TAND.[5,10–12] Therefore, measuring the impact of various manifestations of TSC on functioning is both important and challenging, and could improve care and allow monitoring over time. Moreover, if individuals with TSC have learning difficulties and mental health problems, they may have difficulties indicating their symptoms or healthcare needs, resulting in unknown and hence unmet healthcare needs. This could, in turn, lead to impaired functioning.[10,13,14]

Various outcomes have been measured to assess disease severity in TSC research. Clinical or surrogate outcomes are often narrow in their focus, and it is unclear whether changes are relevant. For instance, although (severity of) epilepsy has been directly related to functioning,[6,15] reduction of seizure frequency does not always lead to improved functioning.[16,17] In addition, what clinicians consider relevant is not identical to what individuals with TSC find important. The International Classification of Functioning and Disability (ICF) is a biopsychosocial model of disability based on an integration of the social and medical models of disability (World Health Organization 2001). The ICF conceptualizes a person's level of functioning

as a dynamic interaction between health conditions, environmental factors, and personal factors.

To get a better understanding of functioning and what is relevant to individuals, a patient-reported outcome measure (PROM) would allow an insight into perceived severity and impact. PROMs are questionnaires that measure how an individual experiences his or her own health.[18–20] They have become important for value-based healthcare and shared-decision making[21] and are increasingly used in practice and scientific research to quantify the severity and impact of the diseases on daily functioning from the perspective of the individual. PROMs enable periodical and quantitative evaluation of symptoms and functioning of the patient population. It can thus be used for monitoring and informing care, and as an outcome measure for trials.[22]

Questionnaires commonly used in TSC trials, such as the Pediatric Quality of Life Inventory (PedsQL™ 4.0)[23] and Short-Form 36 (SF-36),[24] do not include disease-specific symptoms and may not be responsive enough for individuals with TSC.[25,26] In addition, adults with TSC may or may not be able to self-report, and most existing questionnaires for adults are most commonly solely available as self-report. Adult proxy-report questionnaires are often unavailable for the domains of interest. It has been suggested that health problems in TSC are underestimated by excluding the more severely affected individuals, preventing them from early interventions.[27–29] Previous clinical trials that did not demonstrate significant clinical benefits based on parent-reported PROMs as primary outcome measures, such as the Aberrant Behavior Checklist – Irritability subscale,[30] have been considered unsuccessful even when secondary outcome measures, such as visual analog scale ratings of parent-nominated problem behaviors or subscales validated for that specific patient population, indicated positive improvements.[31] This raises questions about whether the intervention was truly ineffective or whether the measurement instrument or mode of administration (proxy-report) was not responsive to therapy or suitable for the population being studied.

Especially now that disease-modifying, and often long-term and expensive therapies are increasingly available, there is an urgent need for a TSC-specific PROM to measure effects of clinical parameters and treatment on disease-specific functioning, in both clinical and research settings. The use of a TSC-specific PROM in clinical trials can provide valuable evidence of the risks and benefits of treatments from a patient perspective which can inform regulatory approvals, clinical guidelines, and health policy, as it captures information that is relevant to the individual with TSC.[32] Therefore, the development of a reliable and valid instrument that measures domains and symptoms relevant to individuals with TSC is a top priority for patient organizations, researchers, and healthcare providers.[5,33,34]

The aim of the current study was to develop and validate a TSC-specific PROM that captures the impact of TSC on physical and mental functions, activity and participation, and social support received by individuals with TSC, using the framework of the ICF (World Health Organization, 2001). The questionnaire is called TSC-PROM and consists of separate versions in English and Dutch for self-report and proxy-report, with the latter being the most suitable option to receive information about the possibly experienced issues for individuals who are unable to report on themselves.

Methods

Standards from the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) were used to develop the questionnaire.[33,34] Development and validation consisted of the following studies: 1) PROM development, 2) content validity, and 3) structural validity, internal consistency, and construct validity (Figure 1).

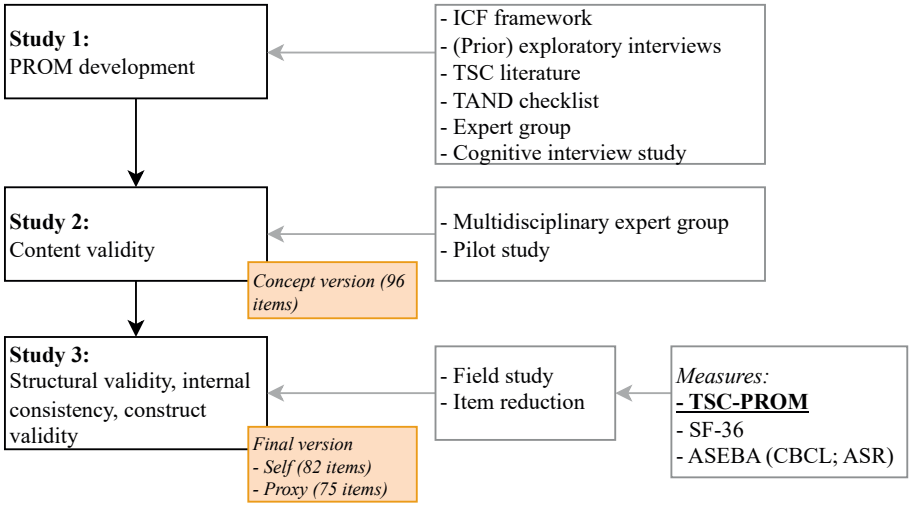


Figure 1. Flowchart of development and validation of the TSC-PROM, according to the standards from the COSMIN.

ASEBA, Achenbach System of Empirically Based Assessment; ASR, Adult Self Report; CBCL, Child Behavior Checklist; ICF, International Classification of Functioning and Disability; PROM, patient-reported outcome measure; SF-36, Short-Form 36; TAND, TSC-associated neuropsychiatric disorders.

Study 1: PROM development

Construct and target population

The construct to be measured is the impact of TSC on physical functions, mental functions, activity and participation, and the social support individuals with TSC receive, using the framework of the ICF (World Health Organization 2001).

The purpose of the TSC-PROM is to evaluate and monitor the impact of TSC on functioning, serving as a tool to facilitate detection and discussion of healthcare needs relevant to individuals with TSC before or during a clinical visit. Two versions of the questionnaire were developed: a self-rated questionnaire for individuals with TSC without ID or a mild ID and a proxy-rated questionnaire for parents and caregivers of adults with TSC who could not complete the questionnaire themselves due to ID severity as assessed by primary caregivers or legal representatives.

The TSC-PROM was developed for all adults (18 years or older) diagnosed with TSC. To fill out the questionnaires, individuals with TSC, parents, or caregivers were English or Dutch-speaking.

Concept elicitation (relevance and comprehensiveness)

Relevant themes were identified by conducting interviews with adults with TSC and caregivers of adults.[10] The TSC-Associated Neuropsychiatric Disorders (TAND) checklist, Lifetime Version (TAND-L) [4] and TSC literature on adult manifestations were reviewed to identify additional themes.[1,10,35–37] The TAND checklist was specifically designed as a screening tool for neuropsychiatric manifestations of TSC, and validated, showing sufficient internal consistency and external validity.[38] Additionally, representatives of patient organizations were asked to identify additional themes. Expert meetings with an expert group representing various disciplines including neurology, psychiatry, psychology, endocrinology, nephrology, ID physician, methodological experts, and representatives of individuals with TSC were held to identify and assess the relevance of the themes until consensus was reached (AvE, AR, MdW, LdG, ET, PJ, PdV, LtH, JvdE).

After identifying relevant themes, the expert group categorized TSC-relevant themes by using the framework of the ICF.[39] The ICF delineates several domains, including the components *health condition, body functions & structure, activity, participation, environmental factors,* and *personal factors.* These components were used to classify the TSC-PROM subscales. The component *body functions & structure* was divided into the domains *1) physical functions* and *2) mental functions*. The ICF components *activity* and *participation* were combined into one domain *3) Activities and participation*. The fourth subdomain *social support* was composed of the ICF component *environmental factors*.

From the identified themes, simple and quantitative questions with response options using 4-point Likert scales were formulated by the expert team. Higher scores indicated better situations. The response options and recall period of the past month were chosen based on the expert opinion. Additionally, visual analog scales (VAS) were included per domain. We also included one question about their health-related Quality of Life (HRQoL).

The preliminary test versions of the questionnaires for individuals with TSC and proxies were reviewed and refined by the expert group. Based on their expertise, topics were added, altered, or removed.

Translation

Two questionnaires were initially developed in the Dutch language in the Netherlands and Belgium. After professional expert translation into English using back and forward translation, a pilot was performed in the United States with English-speaking individuals with TSC and representatives (at least five self-rated and five proxy-rated) in order to broaden the population. The pilot concluded with a discussion with the expert group, and final revisions were made to the questionnaires with a final consensus.

Cognitive interview study (comprehensibility)

A cognitive interview study was performed in all participating languages (Dutch and English) to assess the comprehensibility of the questionnaire. Other than those who participated in the concept elicitation, individuals with TSC from the participating outpatient clinics were asked to provide qualitative feedback on the questionnaire. At least five participants were recruited per type of report (self or proxy) with a definite diagnosis of TSC[2] and a minimum age of 18 years for each participating country (the Netherlands, Belgium, and the United States), with an aimed minimum of 30 participants. We aimed to include participants with different ages, gender, and education to have a sample representing the target population. Individuals with TSC were excluded when an additional genetic disorder to TSC was diagnosed. The cognitive interview study included the 'Think aloud' method[33,40,41] and the 'Retrospective Verbal Probing' technique to assess comprehensibility of the instructions, all items, response options, and recall period.[41]

Study 2: Content validity

Face and content validity was examined by the multidisciplinary expert group and with the abovementioned participants from the cognitive interview study in the Netherlands, Belgium, and the United States; these countries were selected by convenience. Individuals with TSC from the participating outpatient clinics were asked to provide feedback on the digital

questionnaire by completing a feedback form by using the 'Retrospective Verbal Probing' technique,[41] and a topic guide by interviewers who were trained specifically for the study. The feedback form consisted of questions on relevance, comprehensiveness, comprehensibility, and practical issues such as ease of use and lay-out. The TSC-PROM was considered feasible when time to complete was below thirty minutes. Comprehensibility was considered sufficient when at least 75% of the participants agreed on clarity of the instructions, items, formulations, response options, and sequence of items. Also, individuals with TSC were asked whether there were missing, redundant, or unclear items, as well as the most important TSC manifestations, to identify possible missing themes. Cognitive debriefing was performed to refine further and focus the items of the questionnaire and to gain insight into the instrument's practical applicability. Group meetings and interviews were recorded and transcribed verbatim. Two researchers were involved in the analysis. The pilot phase concluded with a final discussion with the expert group about whether each item was relevant for the construct of interest and comprehensiveness of the TSC-PROM. .

Study 3: Structural validity, internal consistency, construct validity

Procedure

E-mails with an invitation link and login code granting access to the questionnaires were sent to participants after answering a question about whether the questionnaire will be filled out by either the individual with TSC themselves or a proxy when (assisted) self-report was not possible, as indicated by the primary caregiver, legal representative or clinician. A proxy who declared to know the individual with TSC well, such as the legal representative or primary caregiver, was allowed to fill out the questionnaires.

Participants

To be eligible for participation in this study, adults (18 years or older) with a definite diagnosis of TSC, molecularly or clinically confirmed according to recent recommendations,[1,2] should be English or Dutch-speaking. Participants were recruited from the outpatient TSC clinics at the University Medical Center of Utrecht (Utrecht, the Netherlands), Erasmus Medical Center (Rotterdam, the Netherlands), the University Hospital of Brussels (Brussels, Belgium), Le Bonheur Children's Hospital Memphis (Memphis,

United States), and Cincinnati Children's Hospital Medical Center (Cincinnati, United States). Additionally, participants were recruited via the Dutch (STSN), Belgian (beTSC), and United States (TSC Alliance) patient organizations.

Measures

In addition to the TSC-PROM, the SF-36,[24] including a proxy-report,[42] and scales assessing emotional and behavioral problems from the Achenbach System of Empirically Based Assessment (ASEBA),[43,44] i.e., the Adult Self Report (ASR), the Child Behavior Checklist (CBCL)/1.5-5 and CBCL/6-18 were used for assessing construct validity. Individuals who were mentally competent to fill out the questionnaires themselves received the ASR. For individuals with TSC who could not complete the questionnaires, caregivers or representatives indicated whether the developmental age was below or above the age of six years old, guiding the distribution of either the CBCL/1.5-5 or CBCL6-18 version. Information on measurement properties of these comparator instruments is provided (see Additional file 1).[24,27,38,42–62]

Item reduction

Item reduction was performed by selection based on frequency (at least 85% with response option 'Not at all'), factor loadings, monotonicity, or local independence unless there was a clinical reason to include the item based on expert opinion. Additionally, items of the proxy version were reduced when frequency of the response options 'Don't know' and 'Not applicable' was at least 30%.

Statistical analyses

Statistical analyses were performed using IBM SPSS Statistics 24 and R. Descriptive statistics were used to characterize demographics, clinical variables, and score distributions of the TSC-PROM. Domain scores were calculated as a percentage of the sum of the items within a domain. For the social support domain, the sum was divided by the number of items filled in other than 'unknown' or 'not applicable' times the number of response options to account for the 'unknown' and 'not applicable' response options. The TSC-PROM total score was the average of the domain percentages, excluding the social support domain. The social support domain was

included as a scale to gather information on the type and quantity of the support someone is receiving, which is an important part of the PROM for the sake of completeness, as this could affect physical functions, mental functions, and activities and participation, but not directly a functioning component. A two-sided significance level of 5% was used.

Structural validity was assessed for the subscales 1) physical functions, 2) mental functions, and 3) activities and participation using a confirmatory factor analysis (CFA) (see Additional file 1).

With regard to internal consistency, Cronbach's alpha was calculated for each TSC-PROM subscale, including the continuous HRQoL VAS. A Cronbach's alpha between 0.70 and 0.95 was considered adequate.[63]

Construct validity was examined by correlating the scores of the TSC-PROM with scores of other instruments that assess the same construct to be measured, also known as convergent validity. Regarding convergent validity, correlations were assessed between the TSC-PROM domain scores and the SF-36 physical component score, mental component score, and the total scores of the ASR, CBCL/1.5-5 or CBCL/6-18. Construct validity was considered sufficient if 75% of the hypotheses were met (see Additional file 1). To assess discriminative validity, analyses were performed using group dichotomization or categorization. A priori hypotheses were defined including 1) individuals with TSC2 pathogenic variants will show lower TSC-PROM scores on the physical domain, mental domain, and TSC-PROM HRQoL VAS compared to individuals with a TSC1 pathogenic variant,[56–58] 2) individuals who reported a drastic life event in the past year will show a lower score on the mental functions domain, 3) individuals with a higher number of involved organ systems will show lower scores on the HRQoL VAS,[27] and 4) individuals with the presence of psychiatric diagnoses will show lower TSC-PROM scores on the mental functions domain, activities and participation domain, and HRQoL VAS (see Additional file 1).[27]

Data management

The questionnaires were digitally distributed using LimeSurvey.[64] As the survey did not allow for missing data, no specific missing item analysis

was necessary. Only the principal investigator (AvE) and researcher (AM) had access to the code for each participant that was solely accessible in the secure network environment of the Erasmus Medical Center. Data were stored in LimeSurvey and exported to R for statistical analyses.

Ethics approval and consent to participate

Medical ethical permission for this study was provided at all participating organizations, with the initiating center the Erasmus University Medical Center Rotterdam in the Netherlands (MEC-2018-1507). Consent forms to participate were provided, and competency to fill in the questionnaire was assessed by the involved healthcare professional. In case the individual was not mentally competent to decide on participation, the legal representative decided on behalf of that person.

Results

Study 1: PROM development

Concept elicitation

Concept elicitation resulted in a draft version of the TSC-PROM for both self and proxy-report (73 items from prior exploratory interviews;[10] supplementary 6 items from TSC literature on adult manifestations; 9 items from the TAND checklist;[4] 11 items from the expert group). After an expert meeting, some items were divided into separate items or combined, resulting in a total of 96 items (24 items within the physical functions domain, 43 items within the mental functions domain, 19 items within the activities and participation domain, and 9 items within the social support domain).

The TSC-PROM starts with 15 questions on demographic and clinical information, including clinical and sociodemographic information (age, sex, nationality, age of TSC diagnosis, genetic testing, organs involved, use of medication, epilepsy, level of functioning, educational level, other diagnoses or health conditions, life events). Visual analog scales from 0 to 100 were included on physical functions, mental functions, the ability to perform daily activities, and satisfaction with social support, and a HRQoL VAS was included. During item development, response options were defined using Likert scales with higher scores indicating overall less impairment. To

illustrate, 'a lot' (1), 'somewhat' (2), 'a little' (3), 'not at all' (4) were response options for items on the physical functions and mental functions domain, such as '*During the past month I was bothered by [e.g. difficulty sleeping, skin abnormalities, seizures]*' and *'During the past month I [e.g. experienced restlessness / insecurity / difficulty in meeting new people, felt anxious, had mood swings, I worried about tumor growth / my financial independence]'.* Response options for the activities and participation domain included 'always' (1), 'often' (2), 'sometimes' (3), 'never' (4) with items such as '*During the past month I was limited in [e.g. learning something new, getting along with people I know well, participating in sport/physical exercise]'.* Response options for the social support domain included 'not at all' (1), 'a little' (2), 'mostly' (3), 'completely' (4), 'not applicable' with items such as '*In the past month I was satisfied with [e.g. the support I received from my family / partner / mental healthcare professionals, how my medication is working]'.*

Cognitive interview study

We recruited eleven participants (five self-rated and six proxy-rated) in the Netherlands, ten in Belgium (five self-rated and five proxy-rated), and ten in the United States (five self-rated and five proxy-rated), with a definite diagnosis of TSC[2] and an average age of 34,43 years (range 18-65 years). The questionnaires were completed for eighteen female participants and thirteen male participants. The level of ID differed from fourteen without ID, six with a mild ID, five with a moderate ID, and six with a severe ID. Based on the feedback received during the interview study, the following adjustments were made:

- An introduction was added for each domain to emphasize the subjective experience of possible complaints and how to deal with structurally present complaints.
- Some questions (mainly regarding the demographic and clinical information) were adjusted and reformulated to abate any confusion and redundant information, and to specify some manifestations, such as frequency of seizures and life events.
- Some items were formulated reversely (e.g. 'I like meeting other people') while the majority was about the burden and complaints, causing confusion. These questions were reformulated.

Study 2: Content validity

Content and face validity of the questionnaire were ensured by involving TSC experts in the field, including individuals with TSC and representatives in focus group interviews and the expert multidisciplinary team. In this way, the instrument's content validity was verified by all major stakeholders. We recruited eleven participants (five self-rated and six proxy-rated) in the Netherlands, ten in Belgium (five self-rated and five proxy-rated), and ten participants in the United States (five self-rated and five proxy-rated).

Feasibility: It took participants 16.53 (±5.00, range 10-30) minutes to complete the questionnaire. Participants preferred a digital version and the lay-out was assessed as clear by 85.7% of the participants and 14.3% somewhat agreed.

Comprehensibility: 81% of the participants found the instructions, the items and the formulations clear and 19% somewhat agreed. One participant indicated difficulties when complaints are always present. Small suggestions were made for clarification. 85.7% of the examples provided were clear and understandable. 66,7% indicated clear response options. Feedback included lack of the response option 'not applicable' in the self-report version and difficulty to estimate the applicability of 'not applicable' or 'do not know' for the proxy-report. 85.7% of the participants agreed on the sequence of items.

Relevance and comprehensiveness: Participants did not indicate other items or complaints and agreed on comprehensiveness, completeness, and relevance. All relevant questions were included, although not all questions were applicable to the different levels of functioning.

Study 3: Structural validity, internal validity, construct validity

E-mails with access to the questionnaires were sent to 210 participants, with a response rate of 78%. In total 163 participants completed the TSC-PROM, of whom 114 participants filled in the complete questionnaire battery (85% of self-reporting participants and 46% of proxy-reporting participants). Six and thirteen self-reporting participants and 27 and 36 proxy-reporting participants did not fill out the SF-36 and ASEBA questionnaires, respectively.

The sociodemographic and demographic and clinical characteristics are presented in Table 1. Seven participants reported other nationalities, including Canadian, Australian, British, Spanish, and Finnish.

Table 1. Sociodemographic and clinical characteristics of participants.

Sociodemographic and clinical characteristics	Self (n=85)		Proxy (n=78)	
		%		%
Age (years)	43.3 (range 18-73)		34.8 (range 18-66)	
Sex				
Female	49	57.6	33	42.3
Male	35	41.2	45	57.7
Other	1	1.2	0	0.0
Nationality				
American	14	16.5	7	9.0
Dutch	61	71.8	63	80.8
Belgium	6	7.1	5	6.4
Other (Canadian, Australian, British, Spanish, Finnish)	4	4.7	3	3.8
Age diagnosis TSC (years)	20.7 (range 0-59)		4.2 (range 0-35)	
Genetic cause *				
TSC1 pathogenic variant	15	17.6	6	7.7
TSC2 pathogenic variant	23	27.1	27	34.6
No pathogenic variant identified	4	4.7	5	6.4
Variant of unknown significance	3	3.5	3	3.8
Result unknown	18	21.2	25	32.1
Not genetically tested or unknown	22	25.9	12	15.4
Organs showing symptoms of TSC (e.g., tubers, tumors, pigment changes)				
None	1	1.2	0	0.0
Brain	60	70.6	72	92.3
Skin	71	83.5	72	92.3
Kidneys	73	85.9	72	92.3
Lungs	29	34.1	11	14.1
Eyes	20	23.5	20	25.6
Heart	21	24.7	32	41.0
Mouth	18	21.2	13	16.7
Other (liver, nails, ovaria, pancreas, uterus, teeth, breast, colon, adrenal, intestines, rectum, ears, nose)	16	18.8	18	23.1

Table 1. continued.

Sociodemographic and clinical characteristics	Self (n=85)		Proxy (n=78)	
		%		%
Use of medication	58	68.2	77	98.7
Anti-seizure drugs	24	28.2	63	80.8
mTOR inhibitors	29	34.1	34	43.6
Anti-hypertensive drugs	11	12.9	14	17.9
Other (e.g., antidepressants, antipsychotics, antihistamines, proton pump inhibitors)	26	30.6	46	59.0
Epilepsy (current or past)	33	38.8	73	93.6
Age first seizure	6.5 (range 0-58)		1.3 (range 0-27)	
Level of intellectual functioning				
Normal intellectual ability	64	75.3	1	1.3
Normal intellectual ability with specific learning disability (dyscalculia, dyslexia)	12	14.1	0	0.0
Mild or moderate ID	9	10.6	31	39.7
Severe or profound ID	0	0.0	46	59.0
Living situation				
Alone	16	18.8	8	10.3
With partner (and/or family)	49	57.6	2	2.6
With my parents (and siblings)	14	16.5	24	30.8
With roommates, friends or others	6	7.1	42	53.8
Support (living)				
Without assistance	71	83.5	0	0.0
Ambulatory professional support (no 24 hour care)	8	9.4	13	16.7
Ambulatory professional support (with 24 hour care)	6	7.1	65	83.3
Other diagnoses *				
Autism spectrum disorders (ASD)	6	7.1	39	50.0
Attention deficit (hyperactivity) disorders (AD(H)D)	3	3.5	8	10.3
Obsessive-compulsive disorders (OCD)	3	3.5	5	6.4
Anxiety disorder	10	11.8	8	10.3
Depressive disorder	14	16.5	7	9.0
Psychotic disorder	0	0.0	6	7.7

Table 1. continued.

Sociodemographic and clinical characteristics	Self (n=85)		Proxy (n=78)	
		%		%
Relation to the individual				
Father			20	25.6
Mother			41	52.6
Sibling			6	10.3
Caretaker			2	6.4
Other			5	5.1

* as reported by the primary caregiver or legal representative who completed the questionnaires.

Item reduction

The TSC-PROM consisting of 96 items and the five visual analog scales was subjected to item reduction by applying the criteria defined in the method section, unless there was a clinical reason to include based on expert opinion. Two items were included based on expert opinion, namely the burden of seizures and kidneys over the past month. After item reduction, the self version contained 82 items and the proxy version 75 items (physical functions domain 18 items and 19 items with an additional item on side effects, mental functions domain 37 items and 28 items, activities and participation domain 13 items and 14 items, social support 13 items, for self-report and proxy-report respectively) (see Additional file 1). Items in the proxy version that relied on internal perception or were difficult to estimate as a proxy were removed, such as '*the individual felt lonely*'.

Structural validity

The mental functions and activities and participation self-report scales displayed sufficient unidimensionality and monotonicity according to the predefined criteria (Table 2). The physical self-report scale and the proxy-report scales did not or only partially satisfy the unidimensionality or monotonicity assumption. Some items within the self-report scale displayed local dependence (residual correlation >0.20; physical functions domain: 4.90%, mental functions domain: 5.03%, activities and participation domain: 5.77%). Within the proxy-report scale, some items showed local dependence (physical functions domain: 8.77%, mental functions domain: 10.46%, activities and participation domain: 9.34%).

Table 2. Unidimensionality and monotonicity of the self and proxy version of the TSC-PROM.

	Self			Proxy		
	Physical functions	*Mental functions*	*Activities and participation*	*Physical functions*	*Mental functions*	*Activities and participation*
Unidimensionality						
CFA						
CFI	0.90	**0.97**	**0.97**	0.88	0.90	**0.96**
TLI	0.89	**0.97**	**0.97**	0.86	0.90	**0.95**
RMSEA	**0.07**	**0.06**	0.11	**0.08**	**0.08**	0.12
Bi-factor model						
ωh	0.59	0.72	0.75	0.53	0.67	0.72
ECV	0.53	**0.60**	**0.64**	0.40	0.53	0.58
Monotonicity						
Hi	>0.10	**>0.31**	**>0.50**	>0.02	>0.06	>0.21
H	0.30	**0.518**	**0.62**	0.27	0.37	0.49

Values in bold indicate acceptable fit. Predefined criteria are provided (see Additional file 1). CFA, confirmatory factor analysis; CFI, comparative fit index; ECV, explained common variance; RMSEA, root mean square error of approximation; TLI, Tucker-Lewis index.

Internal consistency reliability

For the self-report, the corrected item-total correlations ranged from 0.00 to 0.62 (physical functions domain), 0.39 to 0.82 (mental functions domain), 0.56 to 0.84 (activities and participation domain). For the proxy-report, it ranged from 0.00 to 0.72 (physical functions domain), 0.08 to 0.84 (mental functions domain), 0.23 to 0.82 (activities and participation domain). Cronbach's alpha value of the total TSC-PROM score was 0.819 and 0.775 for the self and proxy-report, respectively, which met the threshold criterion range of 0.70-0.95. Cronbach's alpha of each subscale ranged from 0.81 to 0.97 (Table 3).

Table 3. Item-total correlations and internal consistency reliability of the TSC-PROM.

	Item-total correlations		Cronbach's alpha	
	Self-report	*Proxy-report*	*Self-report*	*Proxy-report*
Total TSC-PROM			0.819	0.775
Physical functions domain	0.00-0.62	0.00-0.72	0.806	0.820
Mental functions domain	0.39-0.82	0.08-0.84	0.967	0.924
Activities and participation	0.56-0.84	0.23-0.82	0.938	0.906

Construct validity

All hypotheses regarding construct validity were met for both the self version and proxy version (Table 4).

Table 4. Predefined hypotheses and results regarding construct validity of the TSC-PROM self and proxy version.

Predefined hypotheses	Results			
	Self (n=85)		*Proxy (n=78)*	
Moderately strong correlation between TSC-PROM physical functions domain and SF-36 physical component score	Moderately strong	*r*=0.60; p<0.001	Moderately strong	*r*=0.55, p<0.001
Moderately strong correlation between TSC-PROM mental functions domain and SF-36 mental component score	Strong	*r*=0.83, p<0.001	Moderately strong	*r*=0.53, p<0.001
Moderately strong correlation between TSC-PROM mental functions domain and total ASR or CBCL scores	Strong	*r*=0.87, p<0.001	Moderately strong	*r*=0.61, p<0.001
Weak to moderately strong correlations between TSC-PROM domain score and TSC-PROM HRQoL VAS score				
Physical functions domain	Moderately strong	*r*=0.59, *p*<0.001	Weak	*r*=0.44, *p*<0.001
Mental functions domain	Moderately strong	*r*=0.55, *p*<0.001	Weak	*r*=0.39, *p*<0.001
Activities and participation domain	Moderately strong	*r*=0.64, *p*<0.001	Weak	*r*=0.35, *p*=0.003
Moderately strong correlations between TSC-PROM HRQoL VAS score and TSC-PROM VAS domain scores				
Physical functions domain	Moderately strong	*r*=0.67, *p*<0.001	Moderately strong	*r*=0.57, *p*<0.001
Mental functions domain	Moderately strong	*r*=0.65, *p*<0.001	Moderately strong	*r*=0.61, *p*<0.001
Activities and participation domain	Moderately strong	*r*=0.62, *p*<0.001	Moderately strong	*r*=0.59, *p*<0.001

Weak (r>0.3), moderately strong (r>0.5), and strong (r>0.7) correlations.

Discriminative validity

In the self-report, no significant differences were found between individuals with *TSC1* and *TSC2* pathogenic variants. Furthermore, individuals with TSC who experienced a life event showed a lower score on the TSC-PROM mental functions domain (p=0.018, r=-0.26), patients with a higher number of organ manifestations showed a lower HRQoL VAS score (p=0.021, r=-

0.25), and individuals with TSC with the presence of psychiatric diagnoses showed lower TSC-PROM scores on the mental functions domain (p<0.001, r=-0.47), the activities and participation domain (p<0.001, r=-0.40), and HRQoL VAS score (p=0.040, r=-0.22). In the proxy-report, individuals with *TSC2* pathogenic variants showed a lower TSC-PROM score on the mental functions domain compared to individuals with a *TSC1* pathogenic variant (p=0.012, Cohen's D=0.69). No significant differences were found with regard to the experience of a life event, the number of organ manifestations, and the presence of psychiatric diagnoses.

For the self-report, all of the hypotheses were met, except for the hypotheses regarding the *TSC1* and *TSC2* pathogenic variants, resulting in a total of 63% of hypotheses met. For the proxy-report, one out of eight hypotheses was met (13%), which was the hypothesis regarding the effect of *TSC1* and *TSC2* pathogenic variants on the mental functions domain.

Discussion

The TSC-PROM is the first TSC-specific outcome measure comprehensively addressing all relevant aspects of the ICF model for adults with TSC. It is developed and validated according to the gold standard COSMIN, with versions for proxies of individuals with TSC who are unable to use it themselves (see Additional file 2 and 3). The TSC-PROM may be used in both research and clinical settings to assess physical and mental functions, activity and participation, and social support individuals with TSC receive. To date, the TSC-PROM is available in English and Dutch, but translation into other languages and an accessible digitalized version will allow broader evaluation and application of this TSC-specific PROM.

Psychometrics

Psychometric evaluation shows that the TSC-PROM has sufficient validity and reliability to serve as an instrument to systematically gain insight into the impact of TSC on physical functions, mental functions, and activity and participation, and the social support individuals with TSC receive, and provides a vital addition to current clinical outcomes.

The most important part of the development of the TSC-PROM is content validity[65] which was ensured and verified by all major stakeholders, including individuals with TSC and a broad multidisciplinary team of TSC experts. Some adjustments were made based on the feedback received during the cognitive interview study in the participating countries, and feasibility, comprehensibility, relevance, and comprehensiveness were demonstrated. However, cross-cultural validity has not yet been examined, and cultural adaptations may be necessary when using the TSC-PROM in other countries and languages. Satisfactory results were demonstrated on internal consistency and structural validity. Unidimensionality was satisfied, but there was some overlap between items indicated by local dependencies. This may be explained by the fact that items were divided into clusters with overlap in content of symptoms which often co-exist. Satisfactory results were also demonstrated on construct validity, although not all hypotheses with regard to discriminative validity were met, in particular for the proxy version. These results may reflect the heterogeneity of the TSC population and indicate that function for individuals with TSC is difficult to determine by proxy-reports.[42] Furthermore, higher scores of the TSC-PROM indicating better functioning were observed for self-ratings compared to proxy-ratings ($p<0.001$, $r=-0.50$), perhaps because the proxy-ratings concern individuals who are more affected by the neurological manifestations of TSC, or due to bias of the rating as in other studies proxy-raters often seem to assess functioning as worse.[66–68]

Recommendations for use in the care setting

The TSC-PROM can provide quantitative evaluation of the severity and impact of TSC on various health domains and daily functioning from the patient's perspective. As such, it might be used for monitoring and informing care. The instrument might also serve as a tool to facilitate detection of healthcare needs before or during a clinical visit. Although it is an elaborate questionnaire and it might take some time to complete, it consists of all relevant items. However, not all items or domains are always applicable to individuals with TSC due to the heterogeneity and treatment goals. Therefore, a subdomain could be used as well rather than the whole instrument, although it might still be valuable to use all domains in order to not forget about possible manifestations. It ensures an effective follow-up

and timely referral to appropriate care providers. Until now, assessments of disease severity using clinical rating scales such as the clinical global impression scale omitted patient perspectives about issues of relevance to their health. Additionally, it has been pointed out that perception of the individuals' functioning by clinicians and individuals themselves differ.[69,70] Using the TSC-PROM may improve communication between the individual and clinician and treatment outcomes and facilitate shared-decision making, resulting in increased satisfaction with care.

Recommendations for use in research

The TSC-PROM can bridge the gap between care and interventional research. It can be used as an outcome measure to gain insight into patients' perspective on physical functions, mental functions, activities and participation, and the social support individuals with TSC receive, in observational, epidemiological, longitudinal studies and in interventional trials. It can also relate therapeutic or biomarker findings to self-evaluated functioning. This is important for evaluating novel treatments such as anti-seizure medication, mTOR inhibitors, cannabidiol treatments and eventually more (expensive) targeted therapies such as gene or RNA modification.[2,71] Although a TAND-specific outcome measure is under development,[72] the assessment of all relevant health domains in individuals with TSC has been hampered by the lack of a TSC-specific measure,[9] comparable to several other rare diseases for which disease-specific outcome measures have eventually been developed.[28,73–76]

Thus far, generic instruments have been used with the advantage of allowing comparison between different disease (sub)groups. However, these PROMs often do not include all relevant domains of functioning in TSC, or proxy versions for adults are not available.[10,17] As a result, multiple tools have been used in single trials to measure the full impact. As the TSC-PROM addresses all domains of the ICF framework relevant to individuals with TSC while displaying convergent validity to existing generic instruments (SF-36, ASR, CBCL), it may better capture all important manifestations and aspects that impact the functioning of individuals with TSC than existing instruments.

Strengths, limitations and future directions

The TSC-PROM provides an innovative tool to measure what is relevant to individuals with TSC, taking into account the complexity and heterogeneity of the clinical picture of TSC. It has been developed together with individuals with TSC and according to the gold standard COSMIN, providing high relevancy and good quality. It might serve as an example for future work for heterogeneous and complex disorders where existing instruments are unavailable for proxy-report and the domains of interest.

However, limitations of this study are the sample size and representation of a limited number of countries and languages, as there will be differences between countries and cultures regarding healthcare systems. According to COSMIN criteria, a sample size between 50 and 100 per age group is regarded a good sample size for establishing internal consistency and reliability in a PROM.[34] We aimed for a representative sample size of 200 participants, but a part of the participants did not complete the questionnaire battery. The majority of participants were from the Netherlands, although Belgium, American, Canadian, British, Spanish, and Finnish nationalities were included as well, as we recruited in the Netherlands, Belgium, and the United States without restrictions on nationality. In this study, we started to develop a Dutch and English instrument which was tested in the three participating countries. We have not yet examined the applicability for other countries, neither whether cultural adaptations are needed. Next, the TSC-PROM should be translated into other languages such that all individuals with TSC could benefit regardless of their language, country or culture, ensuring inclusivity.

Future interventional studies should evaluate responsiveness to change, test-retest validity and cross-cultural validity of the TSC-PROM and elaborate on discrepancies in functioning between self-reports and proxy-reports in which both the self and proxy versions are completed for one individual. Also, a shortened version of the TSC-PROM or more advanced psychometric methods such as item response theory (IRT-)based instruments might be developed for individuals with mild ID.[77] Ideally, a generic measure should be developed applicable to all rare genetic neurodevelopmental disorders with appropriate versions for different levels of ID with different symptom

checklists to cover relevant disease-specific aspects, as it is not feasible and desirable to have disorder-specific PROMs for all these thousands of disorders.

Conclusions

The TSC-PROM is the first TSC-specific outcome measure for adults with TSC, which has been developed using the ICF structure covering all relevant aspects of physical functions, mental functions, activities and participation, and social support and with input from individuals with TSC, caregivers, clinicians, as well as literature review and psychometric testing. It appears to have adequate to good psychometric properties of acceptability, reliability, and validity. This TSC-specific PROM provides a unique tool to systematically gain insight into the individuals' experiences and monitor trial and therapy outcome, taking into account the complexity and heterogeneity of the clinical picture of TSC, and empowering TSC clinicians and researchers in the optimal care for adults with TSC.

Acknowledgements

We would like to acknowledge the individuals with TSC and their representatives who participated in the study and Jo Anne Nakagawa, Eva Schoeters, and Jan-Paul Wagenaar on behalf of the patient organizations from the US (TSC Alliance), Belgium (BeTSC), and the Netherlands (Stichting Tubereuze Sclerosis Nederland), respectively. We also want to thank Molly Griffith and Lauren Davis from Cincinnati Children's Hospital Medical Center and Le Bonheur Children's Hospital Memphis who helped with recruitment. Prof. Jan van der Ende is thanked for his methodological contribution to developing the TSC-PROM. Furthermore, we are grateful to the Dutch TSC Foundation (Stichting TSC Fonds) for financially enabling this study. AJ is supported by a Senior Clinical Investigator Fellowship from the Research Foundation Flanders (FWO FKM 1805321N). AvE, LdG, MdW, and LtH are member of European Reference Network (ERN) ITHACA. LdG is affiliated with ENDO-ERN.

Funding

The project was funded by the Dutch TSC Foundation but was not involved in the study.

Competing interests

AvE is part of the scientific advisory board, funded by Jazz Pharmaceuticals, and is an investigator in studies of Jazz Pharmaceuticals. PJdV, MdW, and AJ have provided consultancy to Jazz Pharmaceuticals. LtH works at a department with license for distributing the CBCL in Dutch language areas. The other authors declare that they have no competing interests.

Authors' contributions

AM had a major role in the recruitment and acquisition of data, analysis and interpretation of data, and drafting of the manuscript. ML performed the statistical analyses, provided methodological support, and was a major contributor in drafting the manuscript. LH was involved in the validation of the outcome measure, provided methodological support, and contributed to drafting the manuscript. WdRG had a major role in recruitment, acquisition of data, and drafting the manuscript. PJ was part of the expert group, made substantial contributions to the conception and design of the study, and contributed to drafting the manuscript. AR was part of the expert group, made substantial contributions to the conception and design of the study, and contributed to drafting the manuscript. LtH was part of the expert group, made substantial contributions to the conception and design of the study, and contributed to drafting the manuscript. LdG was part of the expert group, had a major role in recruitment, and contributed to drafting the manuscript. MdW was part of the expert group, made substantial contributions to the conception and design of the study, and contributed to drafting the manuscript. AJ made substantial contributions to the conception and design of the study and contributed to drafting the manuscript. TG had a major role in recruitment and acquisition of data and contributed to drafting the manuscript. JC had a major role in recruitment and acquisition of data and contributed to drafting the manuscript. PdV was part of the expert group, made substantial contributions to the conception and design of the study, and contributed to drafting the manuscript. AvE initiated and designed the study, was part of the expert group, had a major role in the acquisition of data, interpreted the data, and drafted the manuscript. All authors read and approved the final manuscript.

References

1. Northrup H, Krueger DA, Roberds S, Smith K, Sampson J, Korf B, et al. Tuberous sclerosis complex diagnostic criteria update: Recommendations of the 2012 international tuberous sclerosis complex consensus conference. Pediatr Neurol. 2013.
2. Northrup H, Aronow ME, Bebin EM, Bissler J, Darling TN, de Vries PJ, et al. Updated International Tuberous Sclerosis Complex Diagnostic Criteria and Surveillance and Management Recommendations. Pediatr Neurol. 2021;123.
3. Curatolo P, Moavero R, de Vries PJ. Neurological and neuropsychiatric aspects of tuberous sclerosis complex. The Lancet Neurology. 2015.
4. De Vries PJ, Whittemore VH, Leclezio L, Byars AW, Dunn D, Ess KC, et al. Tuberous Sclerosis Associated Neuropsychiatric Disorders (TAND) and the TAND checklist. Pediatr Neurol. 2015.
5. Vanclooster S, Bissell S, van Eeghen AM, Chambers N, De Waele L, Byars AW, et al. The research landscape of tuberous sclerosis complex–associated neuropsychiatric disorders (TAND)—a comprehensive scoping review. Vol. 14, Journal of Neurodevelopmental Disorders. 2022.
6. Vergeer M, de Ranitz-Greven WL, Neary MP, Ionescu-Ittu R, Emond B, Sheng Duh M, et al. Epilepsy, impaired functioning, and quality of life in patients with tuberous sclerosis complex. Epilepsia Open. 2019;4(4).
7. Van Heijst BFC, Geurts HM. Quality of life in autism across the lifespan: A meta-analysis. Autism. 2015.
8. Vranceanu AM, Merker VL, Park ER, Plotkin SR. Quality of life among children and adolescents with neurofibromatosis 1: a systematic review of the literature. Journal of Neuro-Oncology. 2015.
9. de Vries PJ, Franz DN, Curatolo P, Nabbout R, Neary M, Herbst F, et al. Measuring health-related quality of life in Tuberous sclerosis complex - psychometric evaluation of three instruments in individuals with refractory epilepsy. Front Pharmacol. 2018.
10. Both P, ten Holt L, Mous S, Patist J, Rietman A, Dieleman G, et al. Tuberous sclerosis complex: Concerns and needs of patients and parents from the transitional period to adulthood. Epilepsy Behav. 2018.
11. De Vries PJ, Belousova E, Benedik MP, Carter T, Cottin V, Curatolo P, et al. Natural clusters of tuberous sclerosis complex (TSC)-associated neuropsychiatric disorders (TAND): New findings from the TOSCA TAND research project. J Neurodev Disord. 2020;12(1).
12. Rietman AB, van Helden H, Both PH, Taal W, Legerstee JS, van Staa AL, et al. Worries and needs of adults and parents of adults with neurofibromatosis type 1. Am J Med Genet Part A. 2018.
13. Ju YJ, Kim TH, Han KT, Lee HJ, Kim W, Lee SA, et al. Association between unmet healthcare needs and health-related quality of life: A longitudinal study. Eur J Public Health. 2017.
14. van Walsem MR, Howe EI, Ruud GA, Frich JC, Andelic N. Health-related quality of life and unmet healthcare needs in Huntington's disease. Health Qual Life Outcomes. 2017.
15. Capal JK, Bernardino-Cuesta B, Horn PS, Murray D, Byars AW, Bing NM, et al. Influence of seizures on early development in tuberous sclerosis complex. Epilepsy Behav. 2017;70.
16. Boylan LS, Flint LA, Labovitz DL, Jackson SC, Starner K, Devinsky O. Depression but not seizure frequency predicts quality of life in treatment-resistant epilepsy. Neurology. 2004;62(2).
17. Krueger DA, Wilfong AA, Mays M, Talley CM, Agricola K, Tudor C, et al. Long-term treatment of epilepsy with everolimus in tuberous sclerosis. Neurology. 2016;87(23).
18. Frew JW, Davidson M, Murrell DF. Disease-specific health related quality of life patient reported outcome measures in Genodermatoses: A systematic review and critical evaluation. Orphanet Journal of Rare Diseases. 2017.
19. Valderas JM, Kotzeva A, Espallargues M, Guyatt G, Ferrans CE, Halyard MY, et al. The impact of measuring patient-reported outcomes in clinical practice: A systematic review of the literature. Quality of Life Research. 2008.

20. Thompson JJ, Kelly KL, Ritenbaugh C, Hopkins AL, Sims CM, Coons SJ. Developing a patient-centered outcome measure for complementary and alternative medicine therapies II: Refining content validity through cognitive interviews. BMC Complement Altern Med. 2011;11.
21. Gibbons C, Porter I, Gonçalves-Bradley DC, Stoilov S, Ricci-Cabello I, Tsangaris E, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. Vol. 2021, Cochrane Database of Systematic Reviews. 2021.
22. Mercieca-Bebber R, King MT, Calvert MJ, Stockler MR, Friedlander M. The importance of patient-reported outcomes in clinical trials and strategies for future optimization. Patient Relat Outcome Meas. 2018.
23. Varni JW, Seid M, Kurtin PS. PedsQL™ 4.0: Reliability and Validity of the Pediatric Quality of Life Inventory™ Version 4.0 Generic Core Scales in Healthy and Patient Populations. Med Care. 2001;39(8).
24. Ware JE. SF-36 Health Survey update. Vol. 25, Spine. 2000.
25. Bijkerk CJ, De Wit NJ, Muris JWM, Jones RH, Knottnerus JA, Hoes AW. Outcome measures in irritable bowel syndrome: Comparison of psychometric and methodological characteristics. Am J Gastroenterol. 2003;98(1).
26. Lee EH, Kwon O, Hahm KB, Kim WH, Kim J Il, Cheung DY, et al. Irritable bowel syndrome-specific health-related quality of life instrument: Development and psychometric evaluation. Health Qual Life Outcomes. 2016;14(1).
27. Zöllner JP, Conradi N, Sauter M, Knuf M, Knake S, Kurlemann G, et al. Quality of life and its predictors in adults with tuberous sclerosis complex (TSC): a multicentre cohort study from Germany. Neurol Res Pract. 2021;3(1).
28. Zeltner NA, Baumgartner MR, Bondarenko A, Ensenauer R, Karall D, Kölker S, et al. Development and psychometric evaluation of the metabQoL 1.0: A quality of life questionnaire for paediatric patients with intoxication-type inborn errors of metabolism. In: JIMD Reports. 2017.
29. Srikrishna S, Robinson D, Cardozo L, Gonzalez J. Is there a discrepancy between patient and physician quality of life assessment? Neurourol Urodyn. 2009;28(3).
30. Aman MG, Singh NN, Stewart AW, Field CJ. Psychometric characteristics of the aberrant behavior checklist. Am J Ment Defic. 1985.
31. Hessl D, Sansone SM, Berry-Kravis E, Riley K, Widaman KF, Abbeduto L, et al. The NIH Toolbox Cognitive Battery for intellectual disabilities: Three preliminary studies and future directions. J Neurodev Disord. 2016;8(1).
32. Cruz Rivera S, McMullan C, Jones L, Kyte D, Slade A, Calvert M. The impact of patient-reported outcome data from clinical trials: perspectives from international stakeholders. J Patient-Reported Outcomes. 2020;4(1).
33. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: A practical guide. Measurement in Medicine: A Practical Guide. 2011.
34. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. Quality of Life Research. 2018.
35. Franz DN, Bissler JJ, McCormack FX. Tuberous sclerosis complex: Neurological, renal and pulmonary manifestations. Vol. 41, Neuropediatrics. 2010.
36. Hasbani DM, Crino PB. Tuberous sclerosis complex. In: Handbook of Clinical Neurology. 2018.
37. Kingswood C, Bolton P, Crawford P, Harland C, Johnson SR, Sampson JR, et al. The clinical profile of tuberous sclerosis complex (TSC) in the United Kingdom: A retrospective cohort study in the Clinical Practice Research Datalink (CPRD). Eur J Paediatr Neurol. 2016;20(2).
38. Leclezio L, Jansen A, Whittemore VH, De Vries PJ. Pilot validation of the tuberous sclerosis-associated neuropsychiatric disorders (TAND) checklist. Pediatr Neurol. 2015.
39. World Health Organization. Towards a Common Language for Functioning , Disability and Health ICF. International Classification. 2002.
40. Willis GB, Artino AR. What Do Our Respondents Think We're Asking? Using Cognitive Interviewing to Improve Medical Education Surveys. J Grad Med Educ. 2013;5(3).
41. Collins D. Pretesting survey instruments: An overview of cognitive methods. In: Quality of Life Research. 2003.

42. Ediebah DE, Reijneveld JC, Taphoorn MJB, Coens C, Zikos E, Aaronson NK, et al. Impact of neurocognitive deficits on patient–proxy agreement regarding health-related quality of life in low-grade glioma patients. Qual Life Res. 2017;26(4).
43. Achenbach TM, Rescorla L a. Manual for the ASEBA Adult Forms & Profiles. English. 2003.
44. Achenbach TM, Ivanova MY, Rescorla LA. Empirically based assessment and taxonomy of psychopathology for ages 1½–90+ years: Developmental, multi-informant, and multicultural findings. Vol. 79, Comprehensive Psychiatry. 2017.
45. Hatano T, Ohno Y, Imai Y, Moritake J, Endo K, Tamari M, et al. Improved health-related quality of life in patients treated with topical sirolimus for facial angiofibroma associated with tuberous sclerosis complex. Orphanet J Rare Dis. 2020;15(1).
46. Achenbach T. Achenbach System of Empirically Based Assessment (ASEBA): Development, Findings, Theory, and Applications. In: Encyclopedia of Autism Spectrum Disorders. 2009.
47. Fong CY, Ng K, Kong AN, Ong LC, Rithauddin MA, Thong MK, et al. Quality of life of children with tuberous sclerosis complex. Arch Dis Child [Internet]. 2019;104(10):972–8. Available from: 10.1136/archdischild-2018-316394.
48. Ebrahimi-Fakhari D, Hussong J, Flotats-Bastardas M, Ebrahimi-Fakhari D, Zemlin M, von Gontard A, et al. Tuberous Sclerosis Complex Associated Neuropsychiatric Disorders and Parental Stress: Findings from a National, Prospective TSC Surveillance Study. Neuropediatrics. 2019;50(5).
49. Overwater IE, Rietman AB, Mous SE, Bindels-De Heus K, Rizopoulos D, Ten Hoopen LW, et al. A randomized controlled trial with everolimus for IQ and autism in tuberous sclerosis complex. Neurology. 2019.
50. van Eeghen AM, Numis AI, Staley BA, Therrien SE, Thibert RL, Thiele EA. Characterizing sleep disorders of adults with tuberous sclerosis complex: A questionnaire-based study and review. Epilepsy Behav. 2011;20(1).
51. Rosseel Y. Lavaan: An R package for structural equation modeling. J Stat Softw. 2012;48.
52. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, et al. Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care. 2007;45(5 SUPPL. 1).
53. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. MPR-online. 2003;8.
54. Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. J Pers Assess. 2010;92(6).
55. Van Der Ark LA. Mokken scale analysis in R. J Stat Softw. 2007;20(11).
56. Jansen FE, Braams O, Vincken KL, Algra A, Anbeek P, Jennekens-Schinkel A, et al. Overlapping neurologic and cognitive phenotypes in patients with TSC1 or TSC2 mutations. Neurology. 2008;70(12).
57. 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.
58. Willems LM, Schubert-Bast S, Grau J, Hertzberg C, Kurlemann G, Wiemer-Kruel A, et al. Health-related quality of life in children and adolescents with tuberous sclerosis complex and their caregivers: A multicentre cohort study from Germany. Eur J Paediatr Neurol. 2021;35.
59. de Vries PJ, Leclezio L, Gardner-Lubbe S, Krueger D, Sahin M, Sparagana S, et al. Multivariate data analysis identifies natural clusters of Tuberous Sclerosis Complex Associated Neuropsychiatric Disorders (TAND). Orphanet J Rare Dis. 2021;16(1).
60. Ware JE, Kosinski M, Gandek B, Aaronson NK, Apolone G, Bech P, et al. The factor structure of the SF-36 Health Survey in 10 countries: Results from the IQOLA Project. J Clin Epidemiol. 1998;51(11).
61. Reed M, Reaume S, Ferro M. Psychometric Properties of the Short Form-36 (SF-36) in Parents of Children with Mental Illness. Psych [Internet]. 2022;4(2):247–57. Available from: https://doi.org/10.3390/psych4020021.
62. Amin S, Mallick AA, Lux A, O'Callaghan F. Quality of life in patients with Tuberous Sclerosis Complex (TSC). Eur J Paediatr Neurol. 2019;23(6).

63. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1).
64. LimeSurvey [Internet]. [cited 2023 May 17]. Available from: https://www.limesurvey.org/.
65. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010.
66. Burks HB, Des Bordes JKA, Chadha R, Holmes HM, Rianon NJ. Quality of life assessment in older adults with dementia: A systematic review. Vol. 50, Dementia and Geriatric Cognitive Disorders. 2021.
67. Griffiths AW, Smith SJ, Martin A, Meads D, Kelley R, Surr CA. Exploring self-report and proxy-report quality-of-life measures for people living with dementia in care homes. Qual Life Res. 2020;29(2).
68. Alcantara J, Ohm J, Alcantara J. Comparison of pediatric self reports and parent proxy reports utilizing PROMIS: Results from a chiropractic practice-based research network. Complement Ther Clin Pract. 2017;29.
69. Janse AJ, Gemke RJBJ, Uiterwaal CSPM, Van Der Tweel I, Kimpen JLL, Sinnema G. Quality of life: Patients and doctors don't always agree: A meta-analysis. J Clin Epidemiol. 2004;57(7).
70. Lunsky Y, Bramston P. A preliminary study of perceived stress in adults with intellectual disabilities according to self-report and informant ratings. J Intellect Dev Disabil. 2006;31(1).
71. Cheah PS, Prabhakar S, Yellen D, Beauchamp RL, Zhang X, Kasamatsu S, et al. Gene therapy for tuberous sclerosis complex type 2 in a mouse model by delivery of AAV9 encoding a condensed form of tuberin. Sci Adv. 2021;7(2).
72. Heunis TM, Bissell S, Byars AW, Capal JK, Chambers N, Cukier S, et al. Empowering Families Through Technology: A Mobile-Health Project to Reduce the TAND Identification and Treatment Gap (TANDem). Front Psychiatry. 2022;13.
73. Adang LA, Gavazzi F, Jawad AF, Cusack S V., Kopin K, Peer K, et al. Development of a neurologic severity scale for Aicardi Goutières Syndrome. Mol Genet Metab. 2020;130(2).
74. Heaney A, Wilburn J, Rouse M, Langmead S, Blakeley JO, Huson S, et al. The development of the PlexiQoL: A patient-reported outcome measure for adults with neurofibromatosis type 1-associated plexiform neurofibromas. Mol Genet Genomic Med. 2020;8(12).
75. Patterson MC, Lloyd-Price L, Guldberg C, Doll H, Burbridge C, Chladek M, et al. Validation of the 5-domain Niemann-Pick type C Clinical Severity Scale. Orphanet J Rare Dis. 2021;16(1).
76. Kolevzon A, Ventola P, Keary CJ, Heimer G, Neul JL, Adera M, et al. Development of an adapted Clinical Global Impression scale for use in Angelman syndrome. J Neurodev Disord. 2021;13(1).
77. Fries JF, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and promis: Assessment of physical function. J Rheumatol. 2014;41(1).

Additional file 1

Methods

Measures

The TAND checklist was used to identify additional items for the physical functions domain. It was specifically designed by a project team member as a screening tool for neuropsychiatric manifestations of TSC [38]. It addresses developmental milestones, level of functioning, behavioral concerns, psychiatric disorders, intellectual ability, academic skills, psychosocial functioning, and ratings of the impact of the neuropsychiatric symptoms. The TAND checklist has been validated, showing good internal consistency and external validity[38,45].

For construct validity, the Short-Form-36 Health Survey (SF-36)[24] including a proxy-report, and scales assessing emotional and behavioural problems from the Achenbach System of Empirically Based Assessment (ASEBA)[43,44], i.e., the Adult Self Report (ASR), the Child Behavior Checklist (CBCL)/1.5-5 and CBCL/6-18 were used.

The SF-36 is a generic measure with 36 items, organized into eight multi-item scales assessing physical functioning, bodily pain, role limitation due to physical health problems, role limitation due to personal or emotional problems, emotional well-being, social functioning, energy/fatigue, and general health perceptions[46]. A proxy version of the SF-36 was available from a previous study[42]. Summary component scores for physical health (PCS) and mental health (MCS) were calculated, ranging from 0-100, with higher scores indicating better health status. The SF-36 has been commonly used in TSC research and showed evidence of validity and reliability as a measure of HRQoL in parents of children with mental illness[47-49].

The Adult Self Report (ASR) and the Child Behavior Checklist (CBCL) are part of the ASEBA questionnaires, assessing competencies (activities, social, school, and total), emotional state, and behavioural problems of children and adolescents[43,44,50]. Total scores were used. The ASR and CBCL

have been considered valid and reliable instruments and commonly used in TSC research[51-54].

Statistical analyses

Structural validity, which is the degree to which the scores of the TSC-PROM are an adequate reflection of the dimensionality of the construct to be measured, was assessed for the subscales 1) physical functions, 2) mental functions, and 3) activities and participation. A confirmatory factor analysis (CFA) for each subscale with weighted least square mean- and variance-adjusted (WLSMV) estimator was performed to assess unidimensionality using the R-package "lavaan (v0.6-3)"[55,56]. We used the following criteria for an acceptable CFA fit: Scaled Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) values>0.95, a standardized root mean square residual (SRMR) value <0.10, and a root mean square error of approximation (RMSEA) value <0.08[57]. If CFA fit did not meet the criteria[56], a bi-factor model was fit to assess whether unidimensionality was sufficient by assessing if the hierarchical omega (ωh) was >0.80 and the explained common variance (ECV) >0.60[58]. Local independence was assessed by the residual correlation in the CFA model. An item pair was considered to be locally independent if the residual correlation was <0.20[56]. Monotonicity was considered sufficient when the item H values of all items were ≥0.30 and the H value of the entire scale was ≥ 0.50, using Mokken scaling[59].

With regard to internal consistency, which refers to the degree of interrelatedness between items, Cronbach's alpha was calculated for each TSC-PROM subscale including the continuous HRQoL VAS. A Cronbach's alpha between 0.70 and 0.95 was considered adequate.

Construct validity, which refers to the assessment of the construct we aimed to assess, was examined by correlating the scores of the TSC-PROM with scores of other instruments that assess the same construct to be measured. Convergent validity is a subtype of construct validity and refers to the degree to which measures that theoretically should be related are in fact related. Regarding convergent validity, correlations were assessed between the TSC-PROM domain scores and the SF-36 physical component score, mental component score and the total scores of the ASR, CBCL/1.5-

5 or CBCL/6-18. Assumptions were tested, including normality, linearity, homoscedasticity, and presence of significant outliers. A moderately strong correlation (Pearson's or Spearman's r > 0.5) was expected between 1) the TSC-PROM physical functions domain score and the SF-36 physical component score, 2) the TSC-PROM mental functions domain score and the SF-36 mental component score, and 3) the TSC-PROM mental functions domain score and the total ASR score or CBCL scores. Also, it was expected that TSC-PROM domains are associated with TSC-PROM VAS scores (weak correlations). Construct validity was considered sufficient if 75% of the hypotheses were met.

Another subtype of construct validity is discriminative validity, which refers to the degree that measures of constructs that theoretically should not be highly related to each other are in fact not found to be highly correlated to each other. To assess discriminative validity, analyses were performed using group dichotomization or categorization. A priori hypotheses were defined including 1) patients with TSC2 mutations will show lower TSC-PROM scores on the physical domain, mental domain, and TSC-PROM HRQoL VAS compared to patients with a TSC1 mutation[60-62], 2) patients who reported a drastic life event in the past year will show a lower score on the mental functions domain, 3) patients with a higher number of involved organ systems will show lower scores on the HRQoL VAS[27], and 4) patients with the presence of psychiatric diagnoses will show lower TSC-PROM scores on the mental functions domain, activities and participation domain, and HRQoL VAS[27]. Assumptions were tested, including normality and homogeneity of variances. Independent samples t-tests or Mann-Whitney U-tests were performed. Effect sizes were calculated via $r = z/\sqrt{N}$.

Results

Item reduction of the TSC-PROM self-report and proxy-report versions per domain with reasons for exclusion.

	Self	Proxy	*Reason for exclusion*
Physical functions		During the past month the individual was bothered by **dizziness**	- Frequency response options 'Don't know' & 'Not applicable' >30%
	During the past month I was bothered by **problems eating *(f.e. eating too much or too little, eating unusual things)***		- Local independence; This item is associated with the item 'problems with stools (f.e. constipation or diarrhea)'. In clinic, there is generally less focus on stool problems although constipation is a clinically relevant problem, according to the experts.
	During the past month I was bothered by **problems with my hearing or ears *(f.e. infections, hearing loss)***	During the past month the individual was bothered by **problems with hearing or ears *(f.e. infections, hearing loss)***	- Frequency response option 'Not at all' >85% - Low factor loading
	During the past month I was bothered by **heart or vascular problems *(f.e. rhythm abnormalities)***	During the past month the individual was bothered by **heart or vascular problems *(f.e. rhythm abnormalities)***	- Frequency response option 'Not at all' >85%
	During the past month I was bothered by **spasticity *(high muscle tone)***		- Frequency response option 'Not at all' >85%

	Self	Proxy	*Reason for exclusion*
Mental functions		During the past month the individual experienced **problems with certain skills** ***(f.e. arithmetic, reading, writing)***	- Frequency response options 'Don't know' & 'Not applicable' >30%
		During the past month the individual experienced **shyness**	- Frequency response options 'Don't know' & 'Not applicable' >30%
		During the past month the individual experienced **difficulty with self-acceptance**	- Frequency response options 'Don't know' & 'Not applicable' >30%
	During the past month I experienced **difficulty dealing with addictive substances** ***(f.e. alcohol, drugs, gaming)***	During the past month the individual experienced **difficulty dealing with addictive substances** ***(f.e. alcohol, drugs, gaming)***	- Low factor loading - Monotonicity - Additionally for proxy: Frequency response options 'Don't know' & 'Not applicable' >30%
		During the past month the individual felt **unhappy, sad or depressed**	- Frequency response options 'Don't know' & 'Not applicable' >30% - These items rely on internal perception; difficult to estimate as a proxy
		During the past month the individual felt **nervous or stressed**	- Frequency response options 'Don't know' & 'Not applicable' >30% - These items rely on internal perception; difficult to estimate as a proxy
		During the past month the individual felt **anxious or scared**	- Frequency response options 'Don't know' & 'Not applicable' >30% - These items rely on internal perception; difficult to estimate as a proxy
		During the past month the individual felt **lonely**	- Frequency response options 'Don't know' & 'Not applicable' >30% - These items rely on internal perception; difficult to estimate as a proxy
	During the past month I **thought about killing myself (suicide)**	During the past month the individual **talked about killing him/herself (suicide)**	- Frequency response option 'Not at all' >85%
		During the past month the individual **worried a lot**	- Frequency response options 'Don't know' & 'Not applicable' >30%
	During the past month I **saw or heard things that other people did not see or hear** ***(f.e. hallucinations)***		- Frequency response option 'Not at all' >85%

	Self	Proxy	*Reason for exclusion*
Functioning in daily life	During the past month I **was physically aggressive towards others** ***(f.e. throwing things, kicking, hitting)***		- Frequency response option 'Not at all' >85%
	During the past month I **tried to hurt myself**		- Frequency response option 'Not at all' >85%
		During the past month the individual worried about **tumor growth**	- Frequency response options 'Don't know' & 'Not applicable' >30%
	During the past month I worried about **family planning** ***(f.e. passing on TSC)***	During the past month the individual worried about **family planning** ***(f.e. passing on TSC)***	- Low factor loading - Monotonicity - Additionally for proxy: Frequency response options 'Don't know' & 'Not applicable' >30%
		During the past month the individual worried about **money** ***(f.e. due to being unable to work or absences during hospital visits)***	- Frequency response options 'Don't know' & 'Not applicable' >30%
		During the past month the individual worried about **financial independency**	- Frequency response options 'Don't know' & 'Not applicable' >30%
		During the past month the individual worried about **social security** ***(f.e. reimbursement of devices or care)***	- Frequency response options 'Don't know' & 'Not applicable' >30%
		During the past month the individual was limited in **planning and organizing**	- Frequency response options 'Don't know' & 'Not applicable' >30%
	During the past month I was limited in **getting along with strangers**		- Local independence (similar to C8)
	During the past month I was limited in **washing and dressing myself**		- Frequency response option 'Not at all' >85%
	During the past month I was limited in **walking independently**		- Frequency response option 'Not at all' >85%
	During the past month I was limited in **caring for my health** ***(f.e. taking medication)***		- Frequency response option 'Not at all' >85%
		During the past month the individual was limited in **his/her financial independency**	- Frequency response options 'Don't know' & 'Not applicable' >30%
		During the past month the individual was limited in **making use of transportation** ***(f.e. driving a car, riding a bike, taking public transportation)***	- Frequency response options 'Don't know' & 'Not applicable' >30%

Additional file 2

TSC-PROM
Questionnaire for adults with Tuberous Sclerosis Complex (TSC)

Explanation and instructions

What is this questionnaire about?
You are filling in this questionnaire because you have Tuberous Sclerosis Complex. This questionnaire is about the complaints and limitations you experience.

The questionnaire consists of the following domains:
- Baseline information
- Physical functions
- Mental functions
- Activities and participation
- Social support
- Quality of life

How to fill in this questionnaire?
Read the instructions with every section carefully.
Choose the answer that is most appropriate. Don't worry if some questions appear not to apply to you. We have to ask the same questions to everybody.

Explanation and examples
Some questions contain a short explanation. This explains the meaning of the mentioned term.
If f.e. (for example) is used, one or more examples follow. Possibly these examples are not applicable to you, but they may help you understand the question better.

Time
Filling in the questionnaire will take approximately 20 min.

Baseline information

1. What is your sex?
   - ☐ Male
   - ☐ Female
   - ☐ Other

2. What is your age?
   _____ years

3. What is your nationality?
   - ☐ American
   - ☐ Other (please specify): _____

4. At what age were you diagnosed with TSC?
   _____ years

5. (a) Has genetic testing been performed?
   - ☐ I don't know (go to question 6)
   - ☐ No (go to question 6)
   - ☐ Yes

   (b) What were the results?
   - ☐ I don't know
   - ☐ TSC1 mutation
   - ☐ TSC2 mutation
   - ☐ No mutation identified
   - ☐ Mutations found but uncertain if they cause TSC

6. Which organs show, or have shown, symptoms of TSC? For example: tubers, tumors, pigment changes?
   - ☐ None
   - ☐ Skin
   - ☐ Lungs
   - ☐ Heart
   - ☐ Brain
   - ☐ Kidneys
   - ☐ Eyes
   - ☐ Mouth
   - ☐ Other, namely _____

7. Do you use medication?
   - ☐ No
   - ☐ Yes (please list all the medication you use) _____
     _____
     _____

8. (a) Do you have epilepsy, now or in the past?
   - ☐ No (go to question 9)
   - ☐ Yes

   (b) At what age did you have your first seizure?

   _____ years and _____ months
   - ☐ I don't know

   (c) How often do you have seizures?
   - ☐ Daily, approximately _____ per day
   - ☐ Weekly, approximately _____ per week
   - ☐ Monthly, approximately _____ per month
   - ☐ I am seizure free, since _____ years old

   (d) Do you have a vagal nerve stimulator?
   - ☐ No
   - ☐ Yes

   (e) Are you on a ketogenic diet?
   - ☐ No
   - ☐ Yes

9. What is your level of intellectual functioning?
   - ☐ Normal intellectual ability
   - ☐ Normal intellectual ability with specific learning disability (dyscalculia, dyslexia)
   - ☐ Mild or moderate intellectual disability
   - ☐ I don't know

10. What was your last measured IQ or developmental age (if known)?
    - ☐ My IQ was _____ measured on _____(date or year)
    - ☐ My developmental age was (approximately) _____ measured on _____ (date or year)
    - ☐ I don't know

11. What is the highest level of education that you have completed?
    - ☐ None
    - ☐ Preschool/kindergarten
    - ☐ Primary education
    - ☐ Primary education, special education program
    - ☐ Lower secondary education (Middle school or Junior High)
    - ☐ Secondary education, special education program
    - ☐ Upper to post-secondary (Senior High school, 1-year certificate programs)
    - ☐ Academic higher education or doctoral (Bachelor, Master, PhD)

12. What is your current living situation?
- ◻ I live alone, without assistance
- ◻ I live with other people, without assistance
- ◻ I live alone with ambulatory professional support
- ◻ I live with other people and with ambulatory professional support
- ◻ I live in an assisted living facility for people with a disability (no 24 hour care)
- ◻ I live in an assisted living facility for people with a disability (with 24 hour care)

13 Have you ever been diagnosed with any of the following?

	No	Yes	I don't know
Autism spectrum disorder (Autism, ASS, PDD-NOS, Asperger)	◻	◻	◻
Attention deficit hyperactivity disorder (ADD, ADHD)	◻	◻	◻
Obsessive compulsive disorder (OCD)	◻	◻	◻
Anxiety disorder	◻	◻	◻
Depressive disorder	◻	◻	◻
Psychotic disorder (f.e. schizophrenia)	◻	◻	◻

Other diagnoses, namely _____

14. (a) Do you have any other health concerns besides your TSC?
- ◻ No (go to question 15)
- ◻ Yes

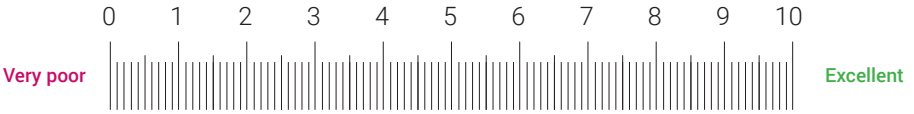
(b) What are these health concerns?
- ◻ High blood pressure (hypertension)
- ◻ Diabetes ('sugar')
- ◻ Thyroid problems
- ◻ Malignant tumor (cancer)
- ◻ Other (please specify): _____

15. (a) In the past year, have you experienced any major life events?
- ◻ No (go to the next section)
- ◻ Yes

(b) What kind of life events?
- ◻ Moving house
- ◻ Change of employment / daytime occupation
- ◻ Severe illness or death of a family member or friend
- ◻ Another major life event, namely _____

Physical functions

In general, how would you rate your overall health? Please put a mark on the ruler below.



Below are complaints or problems related to a person's physical functions that people with or without TSC may experience. Please indicate how much these complaints have troubled you during the last month. If any of the problems are always present, please include them in your estimation of your physical health in the past month.

Reply to each statement by ticking one box per row.

During the past month I was bothered by	A lot	Somewhat	A little	Not at all
1. difficulty sleeping				
2. fatigue				
3. dizziness				
4. problems with my weight *(f.e. unexpected weight loss or weight gain)*				
5. problems with my stomach *(f.e. acid reflux, vomiting, nausea)*				
6. problem with stools *(f.e. constipation or diarrhea)*				
7. problems with my vision or eyes *(f.e. difficulty seeing, squinting)*				
8. speech and/or language problems (*f.e. stuttering, others having difficulty understanding my speech, unintelligible speech*)				
9. problems with my balance *(f.e. difficulty with stability when sitting, standing, or walking)*				
10. problems with my motor skills *(f.e. clumsiness, bad coordination)*				
11. skin abnormalities				
12. inflammation *(f.e. flu, respiratory infection, bladder infection, oral ulcers)*				
13. epileptic insults *(f.e. seizures, staring spells)*				
14. pain				
15. breathing problems *(f.e. shortness of breath, wheezing, coughing)*				
16. problems with my kidneys				
17. fluid retention (*f.e. ankle edema)*				
18. physical problems without a clear cause				
19. During last month I was bothered by side effects from my medication □ No □ Yes (please specify): _____				

Mental functions

In general how was your mental health, including your mood and thinking facilities, during the last month?

Please put a mark on the ruler below.

Below are complaints or problems related to a person's mental functions that people with or without TSC may experience. Please indicate how much these complaints have troubled you during the last month. If any of the problems are always present, please include them in your estimation of your mental health in the past month.

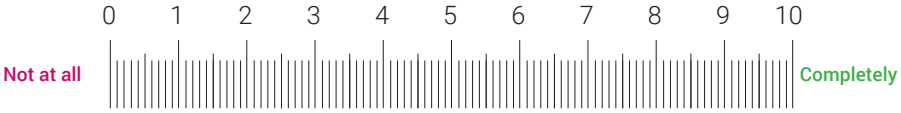
Reply to each statement by ticking one box per row.

During the past month I experienced	A lot	Somewhat	A little	Not at all
1. overactive or hyperactive behavior				
2. restlessness *(f.e. fidgeting or squirming)*				
3. impulsivity *(f.e. doing or saying things without thinking)*				
4. difficulty concentrating or keeping my attention *(f.e. when reading or watching a movie)*				
5. difficulty remembering things				
6. difficulty with orientation in time or place *(f.e. knowing the date, knowing where I am)*				
7. problems with certain skills *(f.e. arithmetic, reading, writing)*				
8. insecurity				
9. shyness				
10. difficulty making eye contact				
11. difficulty relating to peers				
12. difficulty identifying what someone was thinking or feeling				
13. difficulty estimating my own abilities and limitations				
14. difficulty to stand up for myself *(f.e. saying 'no')*				
15. difficulty to accept myself as I am				
16. problems with my kidneys				
17. fluid retention (*f.e. ankle edema)*				
16. difficulty in meeting new people				

17. difficulty with changes in routines				
18. hypersensitivity to sensory stimuli *(f.e. being touched, bright light, busy surroundings)*				
19. the need to repeatedly perform the same actions				
20. stubbornness				
During the past month I felt	A lot	Somewhat	A little	Not at all
21. unhappy, sad or depressed				
22. nervous or stressed				
23. anxious or scared				
24. lonely				
During the past month I	A lot	Somewhat	A little	Not at all
25. had mood swings				
26. had trouble handling stress				
27. panicked easily				
28. worried a lot				
29. couldn't get specific thoughts out of my head				
30. had temper tantrums				
31. was verbally aggressive towards others *(f.e. cursing, scolding)*				
During the past month I worried about	A lot	Somewhat	A little	Not at all
32. tumor growth				
33. epilepsy				
34. side effects of medication				
35. money *(f.e. due to being unable to work or absences during hospital visits)*				
36. my financial independence				
37. my social security *(f.e. reimbursement of devices or care)*				

Activities and participation

During the past month, were you able to do your daily activities (with help as needed)?
Please put a mark on the ruler below.



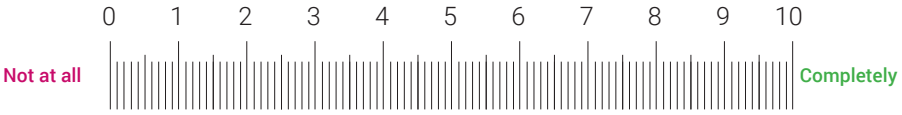
Below, activities are listed that occur in daily life for people with and without TSC. Please score how much you were hindered in performing these activities over the last month.

Reply to each statement by ticking one box per row.

During the past month I was limited in	A lot	Often	Sometimes	never
1. communicating with others				
2. learning something new				
3. planning and organizing				
4. remembering things				
5. doing two things simultaneously *(multi-tasking)*				
6. getting along with people that I know well				
7. building a relationship/making friends				
8. participating in sport/physical exercise				
9. my financial independency				
10. making my own choices *(autonomy)*				
11. managing/planning my own free time				
12. participating in daily activities, work or internship				
13. making use of transportation *(f.e. driving a car, riding a bike, taking public transportation)*				

Social support

In the past month, did you receive the kind of support that you needed?
Please put a mark on the ruler below.



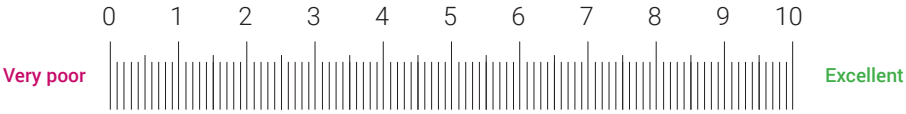
The following statements address how satisfied or dissatisfied you were with different aspects of your life in the past month.

Reply to each statement by ticking one box per row.

In the past month, I was satisfied with the support I received from	No, not at all	A little	Mostly	Completely	Not applicable
1. my family/partner					
2. my friends					
3. patient support groups (patient organisation)					
4. mental healthcare professionals *(f.e. psychiatrist, psychologist, social worker)*					
5. medical professionals *(f.e. doctors, nurses)*					
6. non-medical professionals *(f.e. caretakers)*					
7. from daycare activities, work or internship *(f.e. employer or colleague)*					
In the past month, I was satisfied with	**No, not at all**	**A little**	**Mostly**	**Completely**	**Not applicable**
8. how my medication is working					
9. the home where I live					
10. the availability of information about TSC					
11. my social relationships					
12. my sex life					

Quality of life

The last question asks about quality of life
How would you rate your quality of life over **the past month?**
Please put a mark on the ruler below.



This is the end of the questionnaire.

TSC-PROM
Questionnaire for adults with Tuberous Sclerosis Complex (TSC)

Explanation and instructions

What is this questionnaire about?

You are filling in this questionnaire on behalf of your relative or client with Tuberous Sclerosis Complex.
This questionnaire is about the complaints and limitations he/she experiences.

The questionnaire consists of the following domains:

- Baseline information
- Physical functions
- Mental functions
- Activities and participation
- Social support
- Quality of life

How to fill in this questionnaire?

Read the instructions with every section carefully.
If possible, try and give the answer that you think your relative/client would give.
A statement may be for example: during the last month there was '*worrying or brooding*'.
What matters is not if you were worried, but if your relative/client was worried.
Don't worry if some questions appear not to apply to your relative/client. We have to ask the same questions for everybody.
In the rest of the questionnaire we will call your relative/client 'the individual'.

Explanation and examples

Some questions contain a short explanation. This explains the meaning of the mentioned term.
If f.e. (for example) is used, one or more examples follow. Possibly these examples are not applicable to you, but they may help you understand the question better.

Time

Filling in the questionnaire will take approximately 20 min.

Baseline information

1. What is the individual's sex?
   - ◻ Male
   - ◻ Female
   - ◻ Other

2. What is the individual's age?
   _____ years

3. What is the individual's nationality?
   - ◻ American
   - ◻ Other (please specify): _____

4. What is your relationship to the individual?
   - ◻ Father
   - ◻ Mother
   - ◻ Brother
   - ◻ Sister
   - ◻ Caretaker
   - ◻ Other (please specify): _____

5. At what age was TSC diagnosed?
   _____ years

6. (a) Has genetic testing been performed?
   - ◻ I don't know (go to question 6)
   - ◻ No (go to question 6)
   - ◻ Yes

   (b) What were the results?
   - ◻ I don't know
   - ◻ TSC1 mutation
   - ◻ TSC2 mutation
   - ◻ No mutation identified
   - ◻ Mutations found but uncertain if they cause TSC

7. Which organs show, or have shown, symptoms of TSC? For example: tubers, tumors, pigment changes?
   - ◻ None
   - ◻ Skin
   - ◻ Lungs
   - ◻ Heart
   - ◻ Brain
   - ◻ Kidneys
   - ◻ Eyes
   - ◻ Mouth
   - ◻ Other, namely _____

8. Is the individual using medication?
- ◻ No
- ◻ Yes (please list all medication) _____
  _____
  _____

9. (a) Does the individual have epilepsy, now or in the past?
- ◻ No (go to question 9)
- ◻ Yes

(b) At what age did the first seizure occur?
_____ years and _____ months
- ◻ I don't know

(c) How often do seizures occur?
- ◻ Daily, approximately _____ per day
- ◻ Weekly, approximately _____ per week
- ◻ Monthly, approximately _____ per month
- ◻ The individual is seizure free, since _____ years old

(d) Does the individual have a vagal nerve stimulator?
- ◻ No
- ◻ Yes

(e) Is the individual on a ketogenic diet?
- ◻ No
- ◻ Yes

10. What is the (estimated) level of intellectual functioning of the individual?
- ◻ Normal intellectual ability
- ◻ Normal intellectual ability with specific learning disability (dyscalculia, dyslexia)
- ◻ Mild or moderate intellectual disability
- ◻ Severe or profound intellectual disability
- ◻ I don't know

11. What was the last measured IQ or developmental age of the individual?
- ◻ The IQ was _____ measured on _____ (date or year)
- ◻ The developmental age was (approximately) _____ measured on _____ (date or year)
- ◻ I don't know

12. What is the highest level of education the individual completed?
- ◻ None
- ◻ Preschool/kindergarten
- ◻ Primary education
- ◻ Primary education, special education program
- ◻ Lower secondary education (Middle school or Junior High)
- ◻ Secondary education, special education program
- ◻ Upper to post-secondary (Senior High school, 1-year certificate programs)
- ◻ Academic higher education or doctoral (Bachelor, Master, PhD)
- ◻ I don't know

13. What is the current living situation of the individual?
- □ Alone, without assistance
- □ With other people, without assistance
- □ Alone with ambulatory professional support
- □ With other people and with ambulatory professional support
- □ In an assisted living facility for people with a disability (no 24 hour care)
- □ In an assisted living facility for people with a disability (with 24 hour care)

13 Has the individual ever been diagnosed with any of the following?

	No	Yes	I don't know
Autism spectrum disorder (Autism, ASS, PDD-NOS, Asperger)	□	□	□
Attention deficit hyperactivity disorder (ADD, ADHD)	□	□	□
Obsessive compulsive disorder (OCD)	□	□	□
Anxiety disorder	□	□	□
Depressive disorder	□	□	□
Psychotic disorder (f.e. schizophrenia)	□	□	□

Other diagnoses, namely _____

15. (a) Does the individual has any other health concerns besides TSC?
- □ No (go to question 15)
- □ Yes

(b) What are these health concerns?
- □ High blood pressure (hypertension)
- □ Diabetes ('sugar')
- □ Thyroid problems
- □ Malignant tumor (cancer)
- □ Other (please specify): _____

16. (a) In the past year, has the individual experienced any major life events?
- □ No (go to the next section)
- □ Yes

(b) What kind of life events?
- □ Moving house
- □ Change of employment / daytime occupation
- □ Severe illness or death of a relative or friend
- □ Another major life event, namely _____

Physical functions

In general, how would you rate your overall health? Please put a mark on the ruler below.

Below are complaints or problems related to a person's physical functions that people with or without TSC may experience. Please indicate how much these complaints have troubled the individual during the last month. If any of the problems are always present, please include them in your estimation of the physical health in the past month.

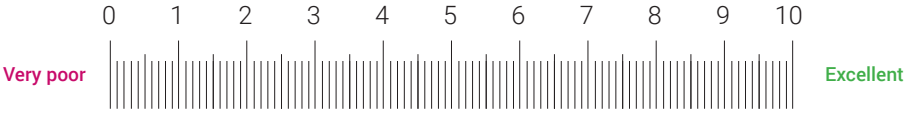
Reply to each statement by ticking one box per row.

During the past month the individual was bothered by	A lot	Somewhat	A little	Not at all	I don't know
1. difficulty sleeping					
2. fatigue					
3. problems eating *(f.e.eating too much or too little, eating unusual things)*					
4. problems with his/her weight *(f.e. unexpected weight loss or weight gain)*					
5. problems with his/her stomach *(f.e. acid reflux, vomiting, nausea)*					
6. problem with his/her stool *(f.e. constipation or diarrhea)*					
7. problems with vision or eyes *(f.e. difficulty seeing, squinting)*					
8. speech and/or language problems (*f.e. stuttering, others having difficulty understanding his/her speech, unintelligible speech*)					
9. problems with the equilibrium *(f.e. balance problems, difficulty with stability when sitting, standing, walking)*					
10. problems with motor skills *(f.e. clumsiness, bad coordination)*					
11. skin abnormalities					
12. inflammation *(f.e. flu, respiratory infection, bladder infection, mouth ulcers)*					
13. epileptic insults *(f.e. seizures, staring spells)*					
14. pain					
15. breathing problems *(f.e. shortness of breath, wheezing, coughing)*					
16. problems with the kidneys					
17. spasticity (*high muscle tone*)					
18. fluid retention (*f.e. ankle edema)*					
19. physical problems without a clear cause					
20. During last month I was bothered by side effects from my medication □ No □ Yes (please specify): _____					

Mental functions

In general how was the mental health of the individual, including your mood and thinking facilities, during the last month?

Please put a mark on the ruler below.



Below are complaints or problems related to a person's mental functions that people with or without TSC may experience. Please indicate how much these complaints have troubled the individual during the last month. If any of the problems are always present, please include them in your estimation of the mental health in the past month.

Reply to each statement by ticking one box per row.

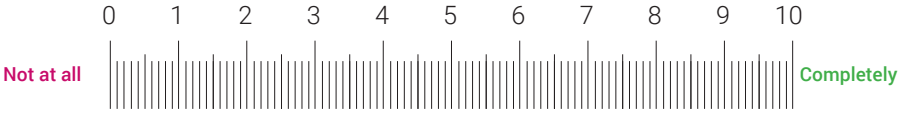
During the past month the individual was bothered by	A lot	Somewhat	A little	Not at all	I don't know	Not applicable
1. overactive or hyperactive behavior						
2. restlessness *(f.e. fidgeting or squirming)*						
3. impulsivity *(f.e. doing or saying things without thinking)*						
4. difficulty concentrating or keeping attention *(f.e. when reading or watching a movie)*						
5. difficulty remembering things						
6. difficulty with orientation in time or place *(f.e. knowing the date, knowing where he/she is)*						
7. insecurity						
8. difficulty making eye contact						
9. difficulty relating to peers						
10. difficulty identifying what someone was thinking or feeling						
11. difficulty estimating his/her own abilities and limitations						
12. difficulty to stand up for him/herself *(f.e. saying 'no')*						
13. difficulty meeting new people						
14. difficulty with changes in routines						
15. hypersensitivity to sensory stimuli *(f.e. being touched, bright light, busy surroundings)*						
16. the need to repeatedly perform the same actions						
17. stubbornness						

During the past month the individual	A lot	Somewhat	A little	Not at all	I don't know	Not applicable
18. had moodswings						
19. had trouble handling stress						
20. panicked easily						
21. saw or heard things that other people did not see or hear *(f.e. hallucinations)*						
22. couldn't get specific thoughts out of his/her head						
23. had temper tantrums						
24. was physically aggressive towards others *(f.e. throwing things, kicking, hitting)*						
25. was verbally aggressive towards others *(f.e. cursing, scolding)*						
26. tried to hurt him/herself						
During the past month the individual worried about	**A lot**	**Somewhat**	**A little**	**Not at all**	**I don't know**	**Not applicable**
27. epilepsy						
28. side effects of medication						

Activities and participation

During the past month, was the individual able to do his/her daily activities (with help as needed)?

Please put a mark on the ruler below.



Below, activities are listed that occur in daily life for people with and without TSC. Please score how much the individual was hindered in performing these activities over the last month.

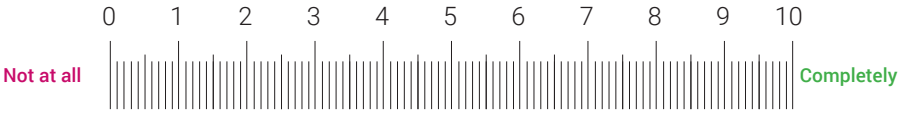
Reply to each statement by ticking one box per row.

During the past month the individual was limited in	Always	Often	Sometimes	Never	I don't know	Not applicable
1. communicating with others						
2. learning something new						
3. remembering things						
4. doing two things simultaneously *(multi-tasking)*						
5. getting along with people that he/she knows well						
6. getting along with strangers						
7. building a relationship/making friends						
8. washing and dressing him/herself						
9. walking independently						
10. participating in sport/physical exercise						
11. caring for his/her health *(f.e. taking medication)*						
12. making his/her own choices *(autonomy)*						
13. managing/planning his/her own free time						
14. participating in daily activities, work or internship						

Social support

In the experience of the individual, did he/she receive the needed support during the last month?

Please put a mark on the ruler below.



The following statements address how satisfied or dissatisfied the individual seems to be with different aspects of his/her life in the past month.

Reply to each statement by ticking one box per row.

In the past month, the individual seems to be satisfied with the support he/she received from	No, not at all	A little	Mostly	Completely	I don't know	Not applicable
1. family/partner						
2. friends						
3. patient support groups (patient organisation)						
4. mental healthcare professionals *(f.e. psychiatrist, psychologist, social worker)*						
5. medical professionals *(f.e. doctors, nurses)*						
6. non-medical professionals *(f.e. caretakers)*						
7. from daycare activities, work or internship *(f.e. employer or colleague)*						
In the past month, the individual seems to be satisfied with	**No, not at all**	**A little**	**Mostly**	**Completely**	**I don't know**	**Not applicable**
8. how the medication is working						
9. the home where he/she lives						
10. the availability of information about TSC						
11. his/her social relationships						
12. his/her sex life						
13. his/her finances						

Quality of life

The last question asks about quality of life
How do you think the individual would rate his/her quality of life over **the past month?**
Please put a mark on the ruler below.

This is the end of the questionnaire.

Additional file 3

TSC-PROM
Vragenlijst voor volwassenen met Tubereuze Sclerose Complex (TSC)

Uitleg en instructies

Waar gaat de vragenlijst over?
U vult deze vragenlijst in omdat u Tubereuze Sclerose Complex heeft.
De vragenlijst gaat over de klachten en beperkingen die u ervaart.

De vragenlijst bestaat uit de volgende delen:
- Basisinformatie
- Lichamelijk functioneren
- Geestelijk functioneren
- Activiteiten en participatie
- Sociale steun
- Kwaliteit van leven

Hoe moet u de vragenlijst invullen?
Lees de instructies bij elk onderdeel zorgvuldig.
Kies het antwoord dat het meest bij u past.
Maakt u zich geen zorgen als sommige vragen niet op u van toepassing zijn. We stellen iedereen dezelfde vragen.

Toelichting en voorbeelden
Bij sommige vragen staat een korte toelichting. Hierin leggen we uit wat we bedoelen met de genoemde term.
Als er 'bijv.' staat, volgen er één of meerdere voorbeelden. Mogelijk zijn deze voorbeelden niet op u van toepassing, maar helpen ze u de vraag beter te begrijpen.

Tijdsduur
Het invullen van de vragenlijst zal ongeveer 20 minuten duren.

Basisinformatie

1. Wat is uw geslacht?
   - ◻ Man
   - ◻ Vrouw
   - ◻ Anders

2. Wat is uw leeftijd?
   _____ jaar _____ maanden

3. Wat is uw nationaliteit?
   - ◻ Nederlands
   - ◻ Belgisch
   - ◻ Anders, namelijk _____

4. Op welke leeftijd werd de diagnose TSC gesteld?
   _____ jaar

5. (a) Is er genetisch onderzoek gedaan?
   - ◻ Weet niet (ga naar vraag 6)
   - ◻ Nee (ga naar vraag 6)
   - ◻ Ja

   (b) Wat was hiervan de uitslag?
   - ◻ Weet niet
   - ◻ TSC1 mutatie
   - ◻ TSC2 mutatie
   - ◻ Geen afwijkingen gevonden
   - ◻ Wel een mutatie gevonden, maar niet zeker of dit de ziekte veroorzaakt

6. In welke organen zijn symptomen van TSC aanwezig, of aanwezig geweest?
   Bijv. tubers, tumoren, pigmentafwijkingen
   - ◻ Geen
   - ◻ Huid
   - ◻ Longen
   - ◻ Hart
   - ◻ Overig, namelijk _____
   - ◻ Hersenen
   - ◻ Nieren
   - ◻ Ogen
   - ◻ Mond

7. Gebruikt u medicatie?
   - ◻ Nee
   - ◻ Ja, namelijk (graag alle huidige medicatie noteren)_____
     _____
     _____

8. (a) Heeft u epilepsie (of in het verleden gehad)?
   - ◻ Nee (ga naar vraag 9)
   - ◻ Ja

   (b) Op welke leeftijd had u uw eerste aanval?
   _____ jaar en _____ maanden

(c) Hoe vaak komen insulten voor?
- ◻ Dagelijks, ongeveer _____ per dag
- ◻ Wekelijks, ongeveer _____ per week
- ◻ Maandelijks, ongeveer _____ per maand
- ◻ Jaarlijks, ongeveer _____ per jaar
- ◻ Ik ben aanvalsvrij, sinds de leeftijd van _____ jaar

(d) Heeft u een nervus vagus stimulator?
- ◻ Nee
- ◻ Ja

(e) Volgt u een ketogeen dieet?
- ◻ Nee
- ◻ Ja

9. Wat is uw niveau van functioneren?
- ◻ Normaal begaafd of bovengemiddeld begaafd
- ◻ Normaal begaafd met specifieke leerproblemen (zoals dyslexie of dyscalculie)
- ◻ Lichte of matige verstandelijke beperking

10. Wat is uw laatst gemeten IQ of ontwikkelingsleeftijd (indien bekend)?
- ◻ Het IQ was _____ gemeten op _____ (datum of jaartal)
- ◻ De ontwikkelingsleeftijd was (ongeveer) _____ gemeten op _____ (datum of jaartal)
- ◻ Weet niet

11. Wat is uw hoogste afgeronde opleiding?

Nederland
- ◻ Geen
- ◻ Basisschool/ Lagere school regulier onderwijs
- ◻ Basisschool/ Lagere school speciaal onderwijs
- ◻ Lager beroepsonderwijs (bijv. huishoudschool, LTS, LEAO, LHNO, praktijkonderwijs)
- ◻ Speciaal voortgezet onderwijs
- ◻ Middelbaar algemeen voortgezet onderwijs (bijv. MAVO, IVO, (M)ULO, VMBO)
- ◻ Middelbare beroepsopleiding (bijv. MBO 2-3, MTS, MEAO, MHNO, INAS)
- ◻ Hoger algemeen en voorbereidend wetenschappelijk onderwijs (bijv. HAVO, VWO, HBS, MMS, Gymnasium, Atheneum)
- ◻ Hoger beroepsonderwijs (bijv. HBO, HTS, HEAO, HHMO)
- ◻ Wetenschappelijk onderwijs (bijv. Bachelor, Master, Master na Master, doctoraat)

België
- ◻ Geen
- ◻ Basisschool/ Lagere school regulier onderwijs
- ◻ Basisschool/ Lagere school buitengewoon onderwijs
- ◻ Lager secundair onderwijs (1ste graad SO)
- ◻ Buitengewoon secundair onderwijs
- ◻ Hoger secundair en post-secundair onderwijs (bijv. ASO, TSO, BSO)
- ◻ Professional bachelor/master
- ◻ Academic master of doctoraat

12. Waar woont u?
- □ Ik woon alleen, zonder begeleiding
- □ Ik woon met anderen, zonder begeleiding
- □ Ik woon alleen en krijg ambulante begeleiding
- □ Ik woon met anderen en krijg ambulante begeleiding
- □ Ik woon in een huis voor mensen met een beperking (géén 24-uurs zorg)
- □ Ik woon in een huis voor mensen met een beperking (24 uurs zorg)

13. Kreeg u, naast de diagnose TSC, ooit één of meerdere van onderstaande diagnoses?

	Nee	Ja	Weet niet
Autisme spectrum stoornis (Autisme, ASS, PDD-NOS, Asperger)	□	□	□
Aandachtstekort-hyperactiviteitstoornis (ADD, ADHD)	□	□	□
Obsessieve-compulsieve stoornis (OCD)	□	□	□
Angststoornis	□	□	□
Depressieve stoornis	□	□	□
Psychotische stoornis (bijv. schizofrenie)	□	□	□

Andere diagnose(s), namelijk _____

14. (a) Heeft u, naast de diagnose TSC, nog andere gezondheidsproblemen?
- □ Nee (ga naar vraag 15)
- □ Ja

(b) Welke gezondheidsproblemen heeft u?
- □ Hoge bloeddruk
- □ Suikerziekte (diabetes)
- □ Schildklieraandoening
- □ Kwaadaardige tumor (kanker)
- □ Anders, namelijk _____

15. (a) Heeft u in het afgelopen jaar een ingrijpende gebeurtenis meegemaakt?
- □ Nee (ga naar het volgende hoofdstuk)
- □ Ja

(b) Wat voor ingrijpends heeft plaatsgevonden?
- □ Verhuizing
- □ Verandering van werk
- □ Ernstige ziekte of overlijden van familie/kennis
- □ Een andere gebeurtenis, namelijk _____

Lichamelijk functioneren

Hoe vindt u over het algemeen uw lichamelijke gezondheid? Plaats een kruisje op de liniaal hieronder.

Hieronder volgen klachten of problemen op het gebied van het lichamelijk functioneren die mensen (met of zonder TSC) kunnen ervaren. Wilt u scoren hoeveel last u heeft gehad van deze klachten gedurende de afgelopen maand? Wanneer de klachten altijd aanwezig zijn, zijn deze dus ook van toepassing op de afgelopen maand.

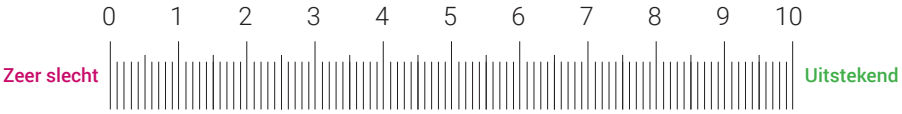
Geef een reactie op elke uitspraak door per rij één hokje aan te vinken.

Gedurende de afgelopen maand had ik last van:	Heel erg	Nogal	Een beetje	Helemaal niet
1. moeite met slapen				
2. moeheid				
3. duizeligheid				
4. problemen met het gewicht *(bijv. onbedoeld afvallen of aankomen)*				
5. maagklachten *(bijv. maagzuur, braken, misselijkheid)*				
6. problemen met de ontlasting *(bijv. verstopping of diarree)*				
7. problemen met zien of de ogen *(bijv. niet goed kunnen zien, scheelzien)*				
8. spraak- en/of taalproblemen *(bijv. stotteren, moeilijk verstaanbaar zijn, moeite met woorden vinden)*				
9. problemen met het evenwicht *(bijv. balansproblemen, moeite met stabiel staan, lopen of zitten)*				
10. problemen van de motoriek *(bijv. onhandigheid, slechte coördinatie, stijfheid)*				
11. huidafwijkingen				
12. ontstekingen of infecties *(bijv. griep, luchtweginfectie, blaasontsteking, aften in de mond)*				
13. epileptische aanvallen *(bijv. trekkingen, staaraanvallen)*				
14. pijn				
15. ademhalingsproblemen *(bijv. kortademigheid, piepen, hoesten)*				
16. nierproblemen				
17. vochtophoping (*bijv. enkeloedeem*)				
18. lichamelijke problemen zonder duidelijke oorzaak				
19. Gedurende de afgelopen maand had ik last van bijwerkingen van de medicatie: □ Nee □ Ja, namelijk _____				

Geestelijk functioneren

Hoe zou u over het algemeen uw geestelijke gezondheid, inclusief uw stemming en denkvermogen, beoordelen?

Plaats een kruisje op de liniaal hieronder..



Hieronder volgen klachten of problemen op het gebied van het geestelijk functioneren die mensen (met of zonder TSC) kunnen ervaren. Wilt u scoren hoeveel last u heeft gehad van deze klachten gedurende de afgelopen maand? Wanneer de klachten altijd aanwezig zijn, zijn deze dus ook van toepassing op de afgelopen maand.

Geef een reactie op elke uitspraak door per rij één hokje aan te vinken.

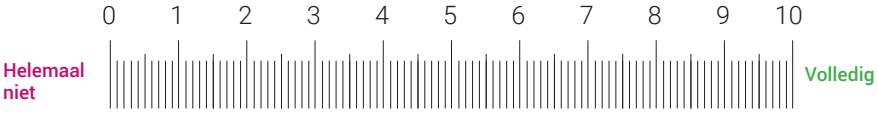
Gedurende de afgelopen maand had ik:	Heel erg	Nogal	Een beetje	Helemaal niet
1. last van overactief of hyperactief gedrag				
2. last van rusteloos of plukkerig gedrag (*bijv. wriemelen of wiebelen*)				
3. last van impulsief gedrag (*bijv. dingen doen of zeggen zonder na te denken*)				
4. moeite met concentreren of lang de aandacht bij iets te houden *(bijv. tijdens lezen of een film kijken)*				
5. moeite met het herinneren van dingen				
6. moeite met het oriënteren in tijd en plaats (*bijv. datum kennen, weten waar u bent)*				
7. moeite met bepaalde vaardigheden *(bijv. rekenen, lezen, schrijven)*				
8. last van onzekerheid				
9. last van verlegenheid				
10. moeite met oogcontact maken				
11. moeite met leeftijdsgenoten om te gaan				
12. moeite met begrijpen wat iemand denkt of voelt				
13. moeite met mijn eigen mogelijkheden en beperkingen in te schatten				
14. moeite met voor mezelf opkomen *(bijv. nee zeggen)*				
15. moeite met mezelf te aanvaarden *(zelfacceptatie)*				
16. moeite met nieuwe mensen ontmoeten				
17. moeite met veranderen van routines				
18. last van overgevoeligheid voor prikkels *(bijv. aanrakingen, fel licht, een drukke omgeving)*				
19. last van het willen herhalen van eenzelfde handeling				
20. last van koppigheid of stuursheid				

Gedurende de afgelopen maand voelde ik me:	Heel erg	Nogal	Een beetje	Helemaal niet
21. ongelukkig, verdrietig of gedeprimeerd				
22. zenuwachtig of gespannen				
23. angstig of bang				
24. eenzaam				
Gedurende de afgelopen maand:	**Heel erg**	**Nogal**	**Een beetje**	**Helemaal niet**
25. kon mijn stemming of gevoel plotseling veranderen				
26. vond ik het moeilijk om te gaan met stress				
27. raakte ik snel in paniek				
28. piekerde ik veel				
29. kon ik bepaalde gedachten moeilijk uit mijn hoofd zetten				
30. had ik driftbuien				
31. was ik verbaal agressief naar anderen *(bijv. vloeken, schelden)*				
Gedurende de afgelopen maand maakte ik me zorgen over:	**Heel erg**	**Nogal**	**Een beetje**	**Helemaal niet**
32. tumorgroei				
33. epilepsie				
34. bijwerkingen van medicatie				
35. geld *(door bijv. niet kunnen werken of verzuim door ziekenhuisbezoeken)*				
36. mijn financiële zelfstandigheid				
37. sociale zekerheid *(bijv. vergoeding van hulpmiddelen of zorg)*				

Activiteiten en participatie

Kon u in de afgelopen maand uw dagelijkse activiteiten ondernemen (eventueel met hulp)?

Plaats een kruisje op de liniaal hieronder.



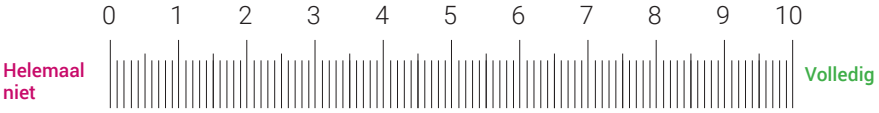
Hieronder volgen activiteiten die plaatsvinden in het dagelijks leven van mensen met en zonder TSC. Wilt u scoren hoeveel belemmeringen u heeft ondervonden gedurende de afgelopen maand met het uitvoeren van deze activiteiten?

Geef een reactie op elke uitspraak door per rij één hokje aan te vinken.

Gedurende de afgelopen maand was ik belemmerd in:	Altijd	Vaak	Soms	Nooit
1. communicatie met anderen				
2. iets nieuws leren				
3. dingen plannen en organiseren				
4. dingen onthouden				
5. twee dingen tegelijkertijd doen (multi-tasking)				
6. omgaan met mensen die ik goed ken				
7. een band opbouwen met iemand of vrienden maken				
8. sporten of aan lichaamsbeweging doen				
9. mijn financiële zelfstandigheid				
10. zelf te kunnen bepalen wat ik doe *(autonomie)*				
11. mijn eigen vrije tijd plannen/indelen				
12. deelname aan werk, stage of dagbesteding				
13. gebruikmaken van vervoermiddelen zoals auto, fiets of openbaar vervoer				
14. participating in daily activities, work or internship				

Sociale steun

Kreeg u in de afgelopen maand het soort steun dat u nodig had?
Plaats een kruisje op de liniaal hieronder.



De volgende uitspraken gaan over hoe tevreden u in de afgelopen maand was met verschillende aspecten van uw leven.

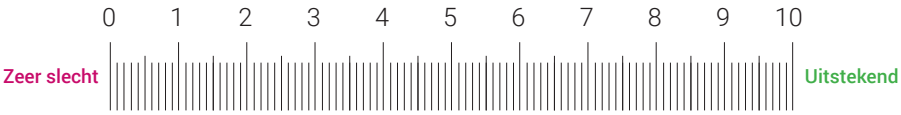
Geef een reactie op elke uitspraak door per rij één hokje aan te vinken.

Ik was in de afgelopen maand tevreden over de steun die ik kreeg van:	Helemaal niet	Enigszins	In grote mate	Volledig	Niet van toepassing
1. mijn familie/partner					
2. mijn vrienden					
3. de patiëntenvereniging					
4. hulpverleners uit de Geestelijke Gezondheidszorg *(bijv. psychiater, psycholoog, maatschappelijk werker)*					
5. medische hulpverleners (*bijv. artsen en verpleegkundigen*)					
6. niet-medische hulpverlening (*bijv. begeleiders)*					
7. werk, stage of dagbesteding *(bijv. collega of werkgever)*					
Ik was in de afgelopen maand tevreden over:	Helemaal niet	Enigszins	In grote mate	Volledig	Niet van toepassing
8. de werking van mijn medicijnen					
9. het huis waar ik woon					
10. de beschikbaarheid van informatie over TSC					
11. mijn sociale relaties					
12. mijn seksuele leven					
13. mijn financiële middelen					

Kwaliteit van leven

De laatste vraag gaat over kwaliteit van leven.

Hoe zou u uw kwaliteit van leven beoordelen **in de afgelopen maand**?
Plaats een kruisje op de liniaal hieronder.



Dit is het einde van de vragenlijst.

TSC-PROM
Vragenlijst voor volwassenen met Tubereuze Sclerose Complex (TSC)

Uitleg en instructies

Waar gaat de vragenlijst over?

U vult deze vragenlijst in voor uw familielid of cliënt met Tubereuze Sclerose Complex.

De vragenlijst gaat over de klachten en beperkingen die hij/zij in het dagelijks leven ervaart.

De vragenlijst bestaat uit de volgende onderdelen:

- Basisinformatie
- Lichamelijk functioneren
- Geestelijk functioneren
- Activiteiten en participatie
- Sociale steun
- Kwaliteit van leven

Hoe moet u de vragenlijst invullen?

Lees de instructies bij elk onderdeel zorgvuldig.

Probeer het antwoord te geven waarvan u denkt dat uw familielid/cliënt zelf zou geven.

Een stelling kan bijvoorbeeld zijn: Gedurende de afgelopen maand was er sprake van *'zich zorgen maken of piekeren'*. Het gaat er dan niet om of u zich zorgen maakte, maar of uw familielid/cliënt zich zorgen maakte.

Maakt u zich geen zorgen als sommige vragen niet op uw familielid/cliënt van toepassing zijn. We moeten iedereen dezelfde vragen stellen.

In de rest van de vragenlijst noemen we 'uw familielid/cliënt' steeds 'betrokkene'.

Toelichting en voorbeelden

Bij sommige vragen staat een korte toelichting. Hierin leggen we uit wat we bedoelen met de genoemde term.

Als er 'Bijv.' staat, volgen er één of meerdere voorbeelden. Mogelijk zijn deze voorbeelden niet op betrokkene van toepassing, maar helpen ze u de vraag beter te begrijpen.

Tijdsduur

Het invullen van de vragenlijst zal ongeveer 20 minuten duren.

Basisinformatie

1. Wat is het geslacht van betrokkene?
   - □ Man
   - □ Vrouw
   - □ Anders

2. Wat is de leeftijd van betrokkene?
   _____ jaar

3. Wat is de nationaliteit van betrokkene?
   - □ Nederlands
   - □ Belgisch
   - □ Anders, namelijk _____

4. Wat is uw relatie tot de betrokkene?
   - □ Vader
   - □ Moeder
   - □ Broer
   - □ Zus
   - □ (Persoonlijk) begeleider
   - □ Anders, namelijk

5. Op welke leeftijd werd de diagnose TSC gesteld?
   _____ jaar

6. (a)Is er genetisch onderzoek gedaan bij betrokkene?
   - □ Weet niet (ga naar vraag 7)
   - □ Nee (ga naar vraag 7)
   - □ Ja

   (b) Wat was hiervan de uitslag?
   - □ Weet niet
   - □ TSC1 mutatie
   - □ TSC2 mutatie
   - □ Geen afwijkingen gevonden
   - □ Wel een mutatie gevonden, maar niet zeker of dit de ziekte veroorzaakt

7. In welke organen zijn symptomen van TSC aanwezig, of aanwezig geweest? Bijv. tubers, tumoren, pigmentafwijkingen
   - □ Geen
   - □ Huid
   - □ Longen
   - □ Hart
   - □ Hersenen
   - □ Nieren
   - □ Ogen
   - □ Mond
   - □ Overig, namelijk _____

8. Gebruikt betrokkene medicatie?
   - ◻ Nee
   - ◻ Ja, namelijk (graag alle huidige medicatie noteren) _____
     _____
     _____

9. (a) Heeft betrokkene epilepsie (of in het verleden gehad)?
   - ◻ Nee
   - ◻ Ja

   (b) Op welke leeftijd had betrokkene de eerste aanval?
   _____ jaar en _____ maanden
   - ◻ Weet niet

   (c) Hoe vaak komen insulten voor?
   - ◻ Dagelijks, ongeveer _____ per dag
   - ◻ Wekelijks, ongeveer _____ per week
   - ◻ Maandelijks, ongeveer _____ per maand
   - ◻ Jaarlijks, ongeveer _____ per jaar
   - ◻ Betrokkene is aanvalsvrij, sinds de leeftijd van _____ jaar

   (d) Heeft betrokkene een nervus vagus stimulator?
   - ◻ Nee
   - ◻ Ja

   (e) Volgt betrokkene een ketogeen dieet?
   - ◻ Nee
   - ◻ Ja

10. Hoe zou u het functioneren van betrokkene inschatten?
    - ◻ Normaal begaafd of bovengemiddeld begaafd
    - ◻ Normaal begaafd met specifieke leerproblemen (zoals dyslexie of dyscalculie)
    - ◻ Milde of matige verstandelijke beperking
    - ◻ Ernstige of diepe verstandelijke beperking

11. Wat is het laatst gemeten IQ of de ontwikkelingsleeftijd (indien bekend) van betrokkene?
    - ◻ Het IQ was _____ gemeten op _____ (datum of jaartal)
    - ◻ De ontwikkelingsleeftijd was _____ gemeten op _____ (datum of jaartal)
    - ◻ Weet niet

12. Wat is de hoogste afgeronde opleiding van betrokkene?

Nederland
- ◻ Geen
- ◻ Basisschool/ Lagere school regulier onderwijs
- ◻ Basisschool/ Lagere school speciaal onderwijs
- ◻ Lager beroepsonderwijs (bijv. huishoudschool, LTS, LEAO, LHNO, praktijkonderwijs)
- ◻ Speciaal voortgezet onderwijs
- ◻ Middelbaar algemeen voortgezet onderwijs (bijv. MAVO, IVO, (M)ULO, VMBO)

- ◻ Middelbare beroepsopleiding (bijv. MBO 2-3, MTS, MEAO, MHNO, INAS)
- ◻ Hoger algemeen en voorbereidend wetenschappelijk onderwijs (bijv. HAVO, VWO, HBS, MMS, Gymnasium, Atheneum)
- ◻ Hoger beroepsonderwijs (bv. HBO, HTS, HEAO, HHMO)
- ◻ Wetenschappelijk onderwijs (WO)

België

- ◻ Geen
- ◻ Basisschool/ Lagere school regulier onderwijs
- ◻ Basisschool/ Lagere school buitengewoon onderwijs
- ◻ Lager secundair onderwijs (1ste graad SO)
- ◻ Buitengewoon secundair onderwijs
- ◻ Hoger secundair en post-secundair onderwijs (ASO, TSO, BSO)
- ◻ Professional Bachelor/Master
- ◻ Academic Master

13. Waar woont betrokkene?

- ◻ Alleen, zonder begeleiding
- ◻ Met anderen, zonder begeleiding
- ◻ Alleen, met ambulante begeleiding
- ◻ Met anderen, met ambulante begeleiding
- ◻ In een huis voor mensen met een beperking (géén 24-uurs zorg)
- ◻ In een huis voor mensen met een beperking (24 uurs zorg)

14. Kreeg betrokkene ooit één of meerdere van onderstaande diagnoses?

	Nee	Ja	Weet niet
Autisme spectrum stoornis (Autisme, ASS, PDD-NOS, Asperger)	◻	◻	◻
Aandachtstekort-hyperactiviteitstoornis (ADD, ADHD)	◻	◻	◻
Obsessieve-compulsieve stoornis (OCD)	◻	◻	◻
Angststoornis	◻	◻	◻
Depressieve stoornis	◻	◻	◻
Psychotische stoornis (bijv. schizofrenie)	◻	◻	◻

Andere diagnose(s), namelijk _____

15. (a) Heeft betrokkene, naast de diagnose TSC, nog andere gezondheidsproblemen?

- ◻ Nee (ga naar vraag 16)
- ◻ Ja

(b) Welke gezondheidsproblemen heeft betrokkene?

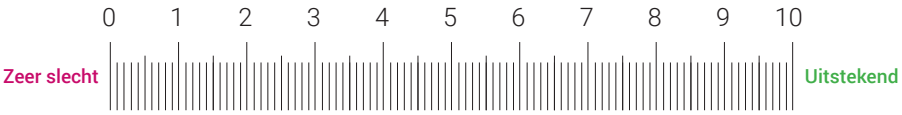
- ◻ Hoge bloeddruk
- ◻ Suikerziekte (diabetes)
- ◻ Schildklieraandoening
- ◻ Kwaadaardige tumor
- ◻ Anders, namelijk _____

16. (a) Heeft betrokkene in het afgelopen jaar een ingrijpende gebeurtenis meegemaakt?
- □ Nee (ga naar het volgende hoofdstuk)
- □ Ja

(b) Wat voor ingrijpends heeft plaatsgevonden?
- □ Verhuizing
- □ Verandering van werk of dagbesteding
- □ Verandering in begeleiding
- □ Problemen met medebewoner of zorg
- □ Ernstige ziekte of overlijden van familie/kennis
- □ Een andere gebeurtenis, namelijk _____

Lichamelijk functioneren

Hoe was de lichamelijke gezondheid van betrokkene in de afgelopen maand?



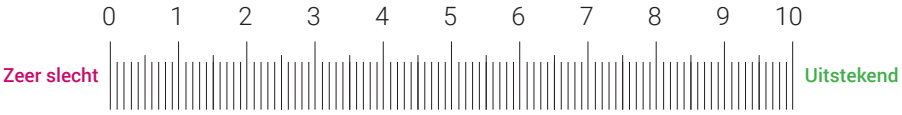
Hieronder volgen klachten of problemen op het gebied van het lichamelijk functioneren die mensen (met of zonder TSC) kunnen ervaren. Wilt u scoren hoeveel last betrokkene heeft gehad van deze klachten gedurende de afgelopen maand? Wanneer de klachten altijd aanwezig zijn, zijn deze dus ook van toepassing op de afgelopen maand

Geef een reactie op elke uitspraak door per rij één hokje aan te vinken.

Gedurende de afgelopen maand had betrokkene last van:	Heel erg	Nogal	Een beetje	Helemaal niet	Weet niet
1. moeite met slapen					
2. moeheid					
3. problemen met eten *(bijv. teveel of te weinig eten, ongewone dingen eten)*					
4. problemen met het gewicht *(bijv. onbedoeld afvallen of aankomen)*					
5. maagklachten *(bijv. maagzuur, braken, misselijkheid)*					
6. problemen met de ontlasting *(bijv. verstopping of diarree)*					
7. problemen met zien of de ogen *(bijv. niet goed kunnen zien, scheelzien)*					
8. spraak- en/of taalproblemen *(bijv. stotteren, moeilijk verstaanbaar zijn, moeite met woorden vinden)*					
9. problemen met het evenwicht *(bijv. balansproblemen, moeite met stabiel staan, lopen of zitten)*					
10. problemen van de motoriek *(bijv. onhandigheid, slechte coördinatie, stijfheid)*					
11. huidafwijkingen					
12. ontstekingen of infecties *(bijv. griep, luchtweginfectie, blaasontsteking, aften in de mond)*					
13. epileptische aanvallen *(bijv. trekkingen, staaraanvallen)*					
14. pijn					
15. ademhalingsproblemen *(bijv. kortademigheid, piepen, hoesten)*					
16. nierproblemen					
17. spasticiteit (*verhoogde spierspanning*)					
18. vochtophoping (*bijv. enkeloedeem*)					
19. lichamelijke problemen zonder duidelijke oorzaak					
20. Had betrokkene gedurende de afgelopen maand last van bijwerkingen van de medicatie: □ Nee □ Ja, namelijk _____					

Geestelijk functioneren

Hoe was over het algemeen de geestelijke gezondheid van betrokkene in de afgelopen maand?



Hieronder volgen klachten of problemen op het gebied van het geestelijk functioneren die mensen (met of zonder TSC) kunnen ervaren. Wilt u scoren hoeveel last betrokkene heeft gehad van deze klachten gedurende de afgelopen maand? Wanneer de klachten altijd aanwezig zijn, zijn deze dus ook van toepassing op de afgelopen maand.

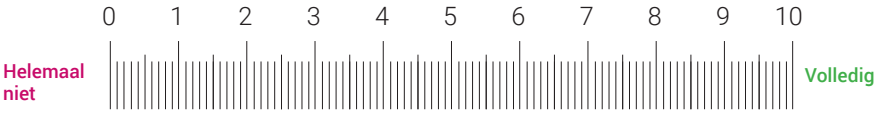
Geef een reactie op elke uitspraak door per rij één hokje aan te vinken.

Gedurende de afgelopen maand had betrokkene:	Heel erg	Nogal	Een beetje	Helemaal niet	Weet niet	Niet van toepassing
1. last van overactief of hyperactief gedrag						
2. last van rusteloos of plukkerig gedrag (*bijv. wriemelen of wiebelen*)						
3. last van impulsief gedrag (*bijv. dingen doen of zeggen zonder na te denken*)						
4. moeite met concentreren of lang de aandacht bij iets te houden *(bijv. tijdens lezen of een film kijken)*						
5. moeite met het herinneren van dingen						
6. moeite met het oriënteren in tijd en plaats (*bijv. datum kennen, weten waar hij/zij is)*						
7. last van onzekerheid						
8. moeite met oogcontact maken						
9. moeite met leeftijdsgenoten om te gaan						
10. moeite met begrijpen wat iemand denkt of voelt						
11. moeite met zijn/haar eigen mogelijkheden en beperkingen in te schatten						
12. moeite met voor zichzelf opkomen *(bijv. nee zeggen*)						
13. moeite met nieuwe mensen ontmoeten						
14. moeite met veranderen van routines						
15. last van overgevoeligheid voor prikkels *(bijv. aanrakingen, fel licht, een drukke omgeving)*						
16. last van het willen herhalen van eenzelfde handeling						
17. last van koppigheid of stuursheid						

Gedurende de afgelopen maand voelde ik me:	Heel erg	Nogal	Een beetje	Helemaal niet	Weet niet	Niet van toepassing
18. had betrokkene stemmingswisselingen						
19. kon betrokkene moeilijk omgaan met stress						
20. raakte betrokkene snel in paniek						
21. zag of hoorde betrokkene dingen die anderen niet zagen *(bijv. hallucinaties)*						
22. kon betrokkene bepaalde gedachten moeilijk uit zijn/haar hoofd zetten						
23. had betrokkene driftbuien						
24. was betrokkene lichamelijk agressief naar anderen *(bijv. dingen gooien, schoppen slaan)*						
25. was betrokkene verbaal agressief naar anderen *(bijv. vloeken, schelden)*						
26. probeerde betrokkene zichzelf te verwonden						
Gedurende de afgelopen maand maakte betrokkene zich zorgen over:	Heel erg	Nogal	Een beetje	Helemaal niet	Weet niet	Niet van toepassing
27. epilepsie						
28. bijwerkingen van medicatie						

Activiteiten en participatie

Kon betrokkene zijn/haar dagelijkse activiteiten ondernemen? (eventueel met hulp)



Hieronder volgen activiteiten die plaatsvinden in het dagelijks leven van mensen met en zonder TSC.

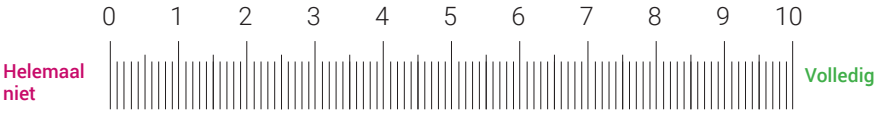
Wilt u scoren hoeveel hinder betrokkene heeft ondervonden gedurende de afgelopen maand in het uitvoeren van deze activiteiten?

Geef een reactie op elke uitspraak door per rij één hokje aan te vinken.

Gedurende de afgelopen maand was betrokkene belemmerd in:	Heel erg	Nogal	Een beetje	Helemaal niet	Weet niet	Niet van toepassing
1. communicatie met anderen						
2. iets nieuws leren						
3. dingen onthouden						
4. twee dingen tegelijkertijd doen (multi-tasking)						
5. omgaan met mensen die hij/zij goed kent						
6. omgaan met onbekenden						
7. een band opbouwen met iemand of vrienden maken						
8. sporten of aan lichaamsbeweging doen						
8. zichzelf wassen en aankleden						
9. zelfstandig lopen						
10. sporten of aan lichaamsbeweging doen						
11. zorgen voor zijn/haar eigen gezondheid *(bijv. medicatie innemen)*						
12. zelf te kunnen bepalen wat hij/zij doet *(autonomie)*						
13. zijn/haar eigen vrije tijd plannen/indelen						
14. deelname aan dagbesteding, stage of werk						

Sociale steun

Kreeg betrokkene in de afgelopen maand het soort steun dat hij/zij nodig had?



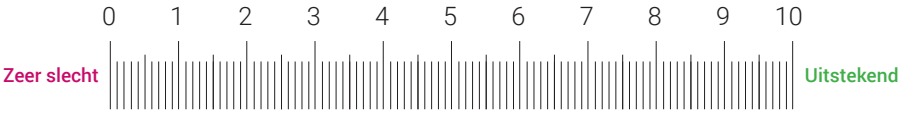
In de volgende vragen wordt gevraagd naar hoe tevreden betrokkene was, of leek te zijn, met de verschillende aspecten van zijn/haar leven.

Geef een reactie op elke uitspraak door per rij één hokje aan te vinken.

Betrokkene was in de afgelopen maand tevreden over de steun van:	Helemaal niet	Enigszins	In grote mate	Volledig	Niet van toepassing
1. familie/partner					
2. vrienden					
3. de patiëntenvereniging					
4. hulpverleners uit de Geestelijke Gezondheidszorg *(bijv. psychiater, psycholoog, maatschappelijk werker)*					
5. medische hulpverleners (*bijv. artsen en verpleegkundigen*)					
6. niet-medische hulpverlening (*bijv. begeleiders)*					
7. dagbesteding, werk of stage *(bijv. collega of werkgever)*					
Betrokkene was in de afgelopen maand tevreden over:	Helemaal niet	Enigszins	In grote mate	Volledig	Niet van toepassing
8. de werking van de medicijnen					
9. het huis waar hij/zij woont					
10. de beschikbaarheid van informatie over TSC					
11. zijn/haar sociale relaties					
12. zijn/haar seksuele leven					
13. zijn/haar financiële middelen					

Kwaliteit van leven

De laatste vraag gaat over kwaliteit van leven.
Hoe zou betrokkene zijn kwaliteit van leven **in de afgelopen maand** inschatten?



Dit is het einde van de vragenlijst.

Chapter 7

Accessibility and feasibility of experience sampling methods for mental health research with people with intellectual disability: scoping of research and stakeholder views

Lianne Bakkum, Célinde Paalman, Annelieke R. Müller,
Agnies M. van Eeghen, Carlo Schuengel

Submitted

Abstract

Background: Experience sampling may be useful for mental health research with people with intellectual disability, but the potential benefits are largely unknown. This multiple-method study investigated the accessibility and feasibility of experience sampling for assessing the mental health of people with intellectual disability.

Method: A scoping review was conducted. Five databases were searched for experience sampling studies involving people with intellectual disability. Seven adults with an intellectual disability tested experience sampling apps with standardised questions about mental health, and were interviewed about their experiences in semi-structured interviews.

Results: Seven studies were included in the scoping review. Two studies investigated feasibility. In the interviews, participants reported on the acceptability, availability, and appropriateness of experience sampling applications.

Conclusions: There are still important gaps in knowledge about acceptability, availability, and appropriateness of experience sampling for people with intellectual disability. Researchers are recommended to tailor experience sampling applications to the needs and preferences of individual users.

Lay summary

- Experience sampling is a method for repeatedly reporting about experiences, such as what participants are doing or feeling at a certain moment.
- Our literature study found that very few studies have used experience sampling for assessing mental health of people with intellectual disability.
- Some adults with mild intellectual disability in our study enjoyed using experience sampling apps and found it useful, and some did not enjoy it.
- Experience sampling methods may be suitable for use in research and clinical practice. It is important for researchers to tailor experience sampling apps to the needs of individual users.

Introduction

In mental health research, retrospective self-report measures such as questionnaires and interviews are widely used for capturing peoples' experiences. These measures often rely on cognitive and communication skills, which can be challenging for people with limited cognitive abilities, such as intellectual disability. Self-report measures are therefore not always inclusive of this population,[1] resulting in their underrepresentation in mental health research. However, the limited available research suggests that people with intellectual disability are as much if not more susceptible to mental health problems.[2] The use of experience sampling may provide an alternative to traditional retrospective methods for assessing mental health, and evidence of the benefits of this method is starting to emerge.[3] In this study, we systematically created an overview of available studies on the feasibility and acceptability of experience sampling methods for assessing mental health of people with intellectual disability. In assessing available research it is important to take the perspectives of people with intellectual disability into account as the primary stakeholders. Our scoping review was therefore informed by questions deemed important by adults with mild intellectual disability who had familiarised themselves with this research method.

Experience sampling is a method for repeated collection of information about people's subjective experiences in real-time, over time, and across different contexts. Initially, this was done using pen and paper.[4] Nowadays, experience sampling studies use digital data collection methods, such as smartphones.[5] Researchers have been using experience sampling methods for assessing mental health from the early 1990s[6] and this field has been rapidly evolving during the last two decades.[7] As summarised by recent reviews, various mental health phenomena have been captured using experience sampling, such as emotion regulation,[8] mood and anxiety,[9] and general well-being[5] in populations without intellectual disability.

Experience sampling studies involving people with intellectual disability have only recently started to emerge, despite the potential benefits of the method for this target population. In experience sampling, participants repeatedly report on their current or very recent inner states, reducing the

memory bias that may be associated with retrospective reports.[10] Indeed, for people with intellectual disability, it may be challenging to recollect and summarise thoughts, mood states, or symptoms exhibited over longer periods.[11] In addition, reports of current states, rather than past states, may be less taxing on people's working memory, which may make this method more suitable for the target group than retrospective reports. Also, recording and analysing fluctuations in mental health over time and within different (social) contexts may give insight into predictors[12] and risk factors for negative moods and symptoms of mental health disorders,[13] as well as the impact of social contexts on mental health.[14] For the target group of people with intellectual disability, this may add to a better understanding of predictors and risk factors for comorbid mental health conditions and evaluation of mental health interventions.

Nonetheless, experience sampling methods may still need to be made accessible for people with intellectual disabilities, just as traditional self-report measures.[1] To conceptualise accessibility, we follow the *Access to Care Framework,*[15] in which five dimensions of accessibility are described: approachability (being aware of the service), acceptability (satisfaction with use), availability (opportunity or capacity to use), affordability (costs), and appropriateness (fit between the service and client needs). The domains of acceptability, availability, and appropriateness may be relevant to take into account when designing and adapting experience sampling methods for mental health research for this target group. Insight into these aspects of accessibility from the viewpoints of people with intellectual disability is crucial.

The aims of this study were: 1) to systematically create an overview of research on the acceptability and feasibility of experience sampling for mental health for people with intellectual disability; 2) to investigate how experience sampling methods for assessing mental health can be made accessible for people with intellectual disability.

Method

Scoping review

Protocol and registration

The protocol for the scoping review was drafted using the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols (PRISMA-P)[16] and was registered on Open Science Framework on 6 December 2022: https://osf.io/qkn8y/?view_only=8007c9f0757b430ab9091a786e2b20df.

Eligibility criteria

Studies were eligible for inclusion if primary (i.e., original) research was reported (any design), participants were children, adolescents or adults with intellectual disability or with genetic neurodevelopmental disorders associated with intellectual disability, and experience sampling methods were digital, self-administered, and used for assessing mental health. Studies using the daily diary method were also included. We used a broad definition of mental health to include psychological and emotional well-being, including quality of life. Only studies in English or Dutch, or other languages that could be understood using Google Translate were included. Excluded from the review were studies targeting elderly with neurodegenerative conditions.

Search strategy

The search date was 2 December 2022, and the search was updated on 12 May 2023. The following databases were sought: IEEE Xplore, Lens, PubMed, Scopus, and Web of Science. Zotero (version 6.0.26) was used for compiling the references. After removing duplicates, the references were uploaded in Rayyan software[17] for title and abstract screening. The final search strategy for all databases can be found on Open Science Framework: https://osf.io/qkn8y/?view_only=8007c9f0757b430ab9091a786e2b20df.

Study selection

Studies were first selected based on the title and abstract by two reviewers, who both screened 50% of the selected items. To reduce bias, 10% of the records were double-screened, after which interrater reliability was calculated. Discrepancies between the reviewers were solved through discussion. After the title and abstract screening, two researchers reviewed

the full-texts of the studies that were selected based on the title and abstract. Both researchers screened 50% of the selected studies, with 10% double-screening. Potential discrepancies were solved through discussion.

Data extraction

Data from the full-text articles were independently extracted by two reviewers with an overlap of two studies. Discrepancies were solved through discussion. We extracted the following bibliographic information: information about the target population, including level of intellectual disability; sample size; mean age of participants; study aim(s); study design; experience sampling design: time-based (interval or random) or event-based; sampling frequency; duration of sampling (e.g., time spent to fill in questionnaire); name of experience sampling application; existing or custom-made experience sampling application; mode of data collection; measures; assessed mental health related outcome(s); compliance and drop-out rates; participants' experiences with experience sampling methods (i.e., acceptability, feasibility). These items were reviewed for necessity and completeness by the research team before extraction.

Synthesis of results

The findings are presented in a narrative synthesis. Information from the selected studies is described in the text and tables.

Interviews with stakeholders

Participants

Seven adults (age 18-64 years) with a mild intellectual disability were included (6 female, 1 male), recruited from 's Heeren Loo, a large organisation providing care for people with intellectual disability in the Netherlands. Six participants were trained as experts-by-experience within 's Heeren Loo. These participants had experience with participating in research projects aimed at improving care for people with intellectual disability. Inclusion criteria were: age ≥ 12 years old, having a smartphone with iOS or Android, being able to independently use a smartphone, and having experience with using smartphone applications.

Design

A qualitative research design was used. Participants each tested two experience sampling applications out of three available applications for three days, with at least seven rest days in between the two testing periods. Four to five prompts (or "beeps") were scheduled at semi-random times throughout the day (depending on the settings of the application). Prompts were scheduled between 8AM and 10PM, with at least 30 minutes between the prompts. The total sampling frequency ranged between 24 and 30 prompts. After each testing period, participants were phoned for a 15-minute debriefing interview. An additional focus group was organised with four participants.

Each participant tested two applications that not only varied in design, but also in response scale. In one of the apps, Likert response scales were used, and in the other app, visual-analogue scales were used. One trained expert-by-experience, who was not part of the study sample, provided advice on the protocol, study design (e.g., feasibility of the number of testing days) and clarity of the questionnaire items. We used different types of response scales to investigate the feasibility and acceptability of these scales in experience sampling questionnaires.

Of the seven participants, five participants tested two apps and took part in two debriefing interviews. Because of personal circumstances, two participants tested one app and took part in one debriefing interview. Thus, in total, 12 debriefing interviews were held. Data saturation was achieved after the twelfth interview. Therefore, no additional participants were recruited.

Materials

Experience sampling method. Three existing experience sampling applications were used: Ethica (2023), m-Path (2023), and Quenza (2023). We selected three applications to explore participants' preferences for design aspects of the applications. All applications allowed for the use of accessibility settings, such as adjusting the font size.

The questionnaire consisted of twelve items from the Positive and Negative Affect Schedule for Children (PANAS-C)[18,19] and four additional questions about social contact. The PANAS-C was used because of the accessible language. The PANAS-C originally consists of 20 items, which were reduced

to twelve to ensure completion of the questionnaire within approximately five minutes. For positive affect, the following items were included: joyful, cheerful, happy, lively, proud, and calm, based on consultation with the expert-by-experience who was involved in the design of the study. For negative affect, we included: miserable, afraid, sad, mad, lonely, and nervous. Examples of questions were: "How cheerful are you feeling right now?" and "How sad are you feeling right now?". The Likert response scale consisted of five response options, ranging from 1 (not at all) to 5 (very much). In the visual-analogue scale option, participants were asked to provide a grade from 0-10 on a line, ranging from 0 (not at all) to 10 (very much).

In the part of the questionnaire about social contact, participants were asked if they were alone at that moment (yes/no), and if so, how they felt being alone (Likert scale ranging from 1 (I don't enjoy it) to 5 (I enjoy it very much); or visual-analogue scale, grade ranging from 0 (not at all enjoying it) to 10 (very much enjoying it)); or if they were just talking, phoning, or texting with someone (yes/no), and if so, with whom (open text field), and how they felt about this (Likert scale ranging from 1 (I don't enjoy it) to 5 (I enjoy it very much); or visual-analogue scale, grade ranging from 0 (not at all enjoying it) to 10 (very much enjoying it)).

Because the purpose of this study was to qualitatively investigate the feasibility of using experience sampling apps in mental health research, data from the questionnaires were not used for answering the research questions and were therefore not analysed. For this reason, the validity and reliability of questionnaire items were not tested.

Debriefing interviews. To examine participants' experiences with using the applications, participants were phoned for a 15-minute semi-structured debriefing interview after testing each of the applications. The first set of interview questions focused on participants' general experiences with the application, such as: experiences with installing; what they liked and did not like about the application; if they experienced any difficulties with using the application; what their opinions were of the number of prompts, the time between prompts, and the number of questions; if they independently answered the questions or if someone helped them; and if they would use the application for longer than three days. The second set of questions

were specific to the experience sampling questions and response scales (Likert response scale or visual-analogue scale), such as: the clarity of the questions; how participants experienced giving a grade to their feelings (visual-analogue scale) or choosing an answer to indicate how they were feeling (Likert response scale); and if participants would have preferred emoticons or other images as response options. The third set of questions directly regarded the feasibility and usefulness of the application, such as: if participants found the application useful for assessing their own mood and feelings; what they would change or improve about the app; and if they would like to use the app in the future. The interviews were audio recorded, transcribed verbatim, and anonymised before analysis.

Focus group. A focus group consisting of four participants was organised to gather additional views on design choices of experience sampling apps. Participants were asked to imagine that they would develop an experience sampling app that would be used by themselves and other individuals with intellectual disability, in the context of mental health research. They were asked what the application would look like, if emoticons or other images should be used, if a reward system (e.g., collecting points) would help completing the questionnaire, what kinds of questions they would include, what the response scales should look like, and what other design choices they would find important.

Procedure

Participants were recruited via the experts-by-experience team from 's Heeren Loo. Information letters and informed consent forms in accessible language were shared with the coaches of the participants, who then shared these materials with them. After informed consent was obtained, individual meetings were scheduled with participants and their coaches to get to know the researchers, to allow participants to ask questions about the study, and to schedule the testing days and debriefing interviews. After the first meeting, participants installed the applications on their smartphone, independently or with help from their coach. Each participant tested two applications for three days, with at least seven rest days in between the two testing periods. During the testing days, participants were prompted five times to complete a five-minute questionnaire. The fifteen-minute debriefing interviews took place as soon as possible after the last testing day, to reduce memory bias.

The focus group was organised after four participants completed testing the second application. Participants received compensation in the form of paying the costs for a group activity. This study has been approved by the ethical review board of the Amsterdam UMC in the Netherlands.

Data analysis

The interview transcripts were coded using deductive thematic analysis by two researchers, following the steps described by Braun and Clarke (2006).[20] The themes were predefined based on three domains of the Care Framework described by Levesque et al. (2013): acceptability (satisfaction with use of the service), availability (opportunity or capacity to use the service), and appropriateness (fit between the service and client needs).

The analysis steps were as follows:[20] 1) both researchers read the transcripts to become familiar with the data; 2) both researchers independently coded two of the same interviews; 3) differences in coding were discussed in a meeting, after consensus was reached about the codes and fit with the themes; 4) the researchers coded a third interview together and discussed the codes and fit with the themes, initial sub-themes were created; 5) both researchers independently coded another two of the same interviews; 6) differences in coding were discussed in a meeting, after consensus was reached about the codes and fit with the main and sub-themes; 7) the codes, main and sub-themes were discussed with the research team, after which consensus was reached about the themes and sub-themes; 8) the remaining interviews were divided for coding between the two researchers. After the twelfth interview, data saturation was discussed. The same themes and sub-themes were used for coding the focus group transcript. ATLAS.ti Mac (Version 23.0.1) was used for analysing the data.

Results

Scoping review

Selection and characteristics of sources of evidence

A total of 971 studies were screened on relevance based on the title and abstract. Two reviewers both screened 50% of selected studies, with an overlap of 10% (agreement: 96%). Discrepancies were solved through

discussion. Sixteen studies were initially screened based on the full-text, of which five studies fit the inclusion criteria.[3,21–24] The search was updated towards the end of the scoping review, which yielded one additional study that fit the criteria.[25] Another eligible study was found using snowballing.[26] In total, seven studies were included in the scoping review.
The PRISMA flow chart is presented in Figure 1.

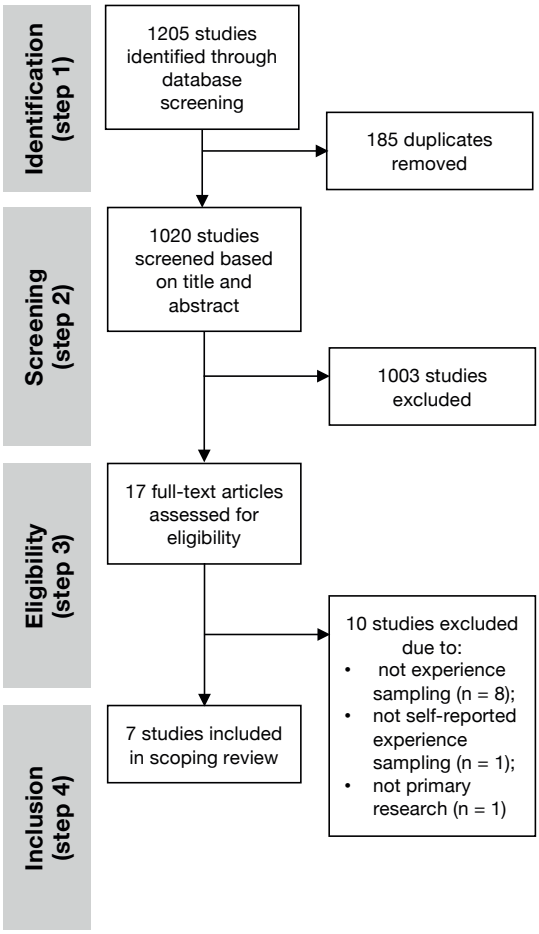


Figure 1. PRISMA flow chart.

Table 1 presents the characteristics of the included studies. The studies were published between 2015 and 2023. In four of the studies,[21–24] people with 22q11.2 deletion syndrome were the target population, one study was focused on acquired brain injury,[27] and in three studies, the target population

Table 1. Characteristics of Included Studies.

Authors	Target population	Sample size	Age of participants (years)	Experience sampling design	Sampling frequency	Duration of sampling
Feller et al. (2023)	22q11.2 deletion syndrome, autism spectrum disorder without intellectual disability	*N* = 107 (*n* = 32 with 22q11.2 deletion syndrome)	*M* = 19.19, *SD* = 4.67	Time-based, semi-random	8 times/ day during 6 consecutive days	Unlimited time, sessions <15 mins were kept in the analysis
Feller et al. (2021)	22q11.2 deletion syndrome	*N* = 86 (*n* = 37 with 22q11.2 deletion syndrome)	*M* = 18.32, *SD* = 3.976	Time-based, semi-random	8 times/ day during 6 consecutive days	2 min 40 s ± 1 min 34 s
Gosens et al. (2023)	Mild intellectual disability and borderline intellectual functioning	*N* = 12	Range: 18-35	Time-based, timing determined in consultation with participants	1 time/ day, study duration between 5-11 months	Participants had 230-720 mins to complete the diary
Hulsmans et al. (2023)	Mild intellectual disability and borderline intellectual functioning	*N* = 50	*M* = 21.4, *SD* = 5.1	Time-based, timing determined in consultation with participants	1 time/day during 60 consecutive days	Participants' reports: completing took between <1 to 8 min (median: 2 min).
Hyde et al. (2021)	22q11.2 deletion syndrome	*N* = 50 (*n* = 29 with 22q11.2 deletion syndrome)	*M* = 33.7 years, *SD* = 9.80	Time-based, random	10 times/ day during 6 consecutive days (affect items)	Not reported
Price et al. (2015)	Acquired brain injury	*N* = 1	26	Unclear	Daily	10 minutes
Schneider et al. (2020)	22q11.2 deletion syndrome	*N* = 55 (*n* = 31 with 22q11.2 deletion syndrome)	*M* = 34.11, *SD* = 9.81	Time-based, semi-random	10 times/ day during 6 consecutive days	Not reported
Wilson et al. (2020)	Mild-to-moderate intellectual disability	*N* = 19	*M* = 27.2, *SD* = 7.7	Time-based, random	7 times/ day during 7 consecutive days	Not reported

Number of questionnaire items	Experience sampling application	Existing or custom-made application	Mode of data collection	Mental health related outcomes	Instruments
33-38	RealLife Exp	Existing application	Mobile phone	Positive affect, negative affect, experience of aloneness, experience of social interaction	Self-designed questionnaire
12	RealLife Exp	Existing application	Mobile phone	Positive affect, negative affect, psychotic experiences	Self-designed questionnaire
2-3	Ethica	Existing application	Mobile phone	Frequency and quantity of substance abuse	Self-designed questionnaire
8	Ethica	Existing application	Mobile phone	Internalising symptoms (anxiety sensitivity, negative thinking); externalising symptoms (impulsivity, sensation seeking)	Self-designed questionnaire
9 (affect items)	PsyMate	Existing application	PsyMate device	Positive affect, negative affect	Self-designed questionnaire
	Not reported	Custom-made application	Custom-made device	Fatigue	Mental Fatigue Scale
				Cognitive attention and working memory	Spatial Span Test from WAIS
				Fatigue measured by reaction time	Psychomotor Vigilance Test
				Cognitive throughput	Serial Addition Subtraction Task
19	PsyMate	Existing application	PsyMate device	Positive affect, negative affect, psychotic experiences, social stress, alone stress, activity-related stress, event-related stress	Self-designed questionnaire
12	mEMA	Existing application	Mobile device (phone or tablet)	External and internal aspects of experiences, including quality of experience and emotions	Self-designed questionnaire

were people with mild intellectual disability[3] or mild intellectual disability/ borderline intellectual functioning.[25,26] Sample sizes of the included studies ranged between 1-107 participants.

Synthesis of results

The study aims and findings regarding the compliance, drop-out rates, and acceptability and feasibility of experience sampling are reported in Table 2.

In two studies,[3,25] the acceptability and feasibility of experience sampling were the target outcomes. In the other studies, experience sampling was used to investigate social and/or psychological outcomes. Compliance rates were reported by five studies[3,22,24–26] and ranged between 34%[3] to 95%.[22] Feller et al. (2021; 2023) reported that the researchers sent personalised messages to the participants every other day to encourage them and to verify study compliance, and that the researchers were available for support throughout the experience sampling procedure. The authors suggested that the high compliance in their study may be due to the guidance of the researchers during the study period. In four of the included studies,[21–24] participants were excluded from the study because compliance was below 33%.

Findings regarding the acceptability and feasibility of experience sampling were reported by three studies.[3,24,25] Hulsmans et al. (2023) concluded that daily experience sampling is feasible for adolescents and young adults with mild intellectual disability and borderline intellectual functioning, based on compliance (70%) and drop-out (26%) rates. Based on interviews, they further concluded that experience sampling is acceptable for this population. The findings suggest that using one's own device (instead of a device shared by a group of participants) for completing the diary is important for compliance with the study protocol. Reasons for non-compliance included, among other reasons, forgetting to fill in the diaries, being in the company of other people, and experiencing high levels of stress. More reminders by care professionals and the application to fill in the diary would have helped participants to increase their compliance. The study duration of 60 days and one prompt per day were appropriate for most participants (69% and 73% of participants, respectively). Further, participants reported gains from participating, the most prominent ones were increased self-

awareness and receiving a gift-card. Drawbacks from participating were, among others, technical difficulties with the app or device, and irrelevance of the questionnaire items. Around half of the participants did not report any drawbacks.

Based on interviews with adults with mild to moderate intellectual disability, Wilson and colleagues (2020) suggest that experience sampling is feasible for this population. Participants enjoyed taking part in the study and would be willing to participate in similar studies in the future. Reasons for non-compliance included being elsewhere with limited access to the device (e.g., at work), leaving the device somewhere else, not hearing the notification beeps, or technical difficulties with the device. In addition, participants and parents of participants provided recommendations for improving the feasibility of experience sampling applications, including the use of longer and louder prompts, use of specific language to avoid confusion about the questions, and include pictographs to aid understanding of the questions.

In the study by Schneider et al. (2020), the researchers found that a subgroup of participants required additional monitoring during the study, but no further details were reported.

Interview findings

Seven adults with mild intellectual disability tested existing experience sampling applications for three consecutive days, receiving four to five prompts per day, and were interviewed afterwards about their experiences. An additional focus group was held with four participants. The interviews were coded using thematic analysis, following Braun and Clarke (2006). Three main themes were predefined based on three domains of access to care as described by Levesque et al. (2013): acceptability (sub-themes: value and usefulness, and practical feasibility), availability (sub-themes: installation and registration, language and text, and response scales), and appropriateness (sub-themes: adapting the experience sampling design and the questionnaire response scales to the needs of users). Table 3 presents an overview of the themes and sub-themes, including examples of codes.

Table 2. Aims of Included Studies and Findings Related to the Feasibility of Experience Sampling.

Authors	Study aim(s)	Compliance and drop-out	Acceptability and feasibility of ESM
Feller et al. (2023)	To characterise social functioning in the daily-life of adolescents and young adults with 22q11.2 deletion syndrome and autism spectrum disorder.	One participant (3%) with 22q11.2 deletion syndrome was excluded because of a compliance of <33%.	Not reported
Feller et al. (2021)	To explore psychotic experiences and their association with affective states in the daily-life of adolescents and young adults with 22q11.2 deletion syndrome using experience sampling, and to examine how these manifestations relate to a gold standard semi-structured assessment of psychotic experiences.	Two participants (5%) with 22q11.2 deletion syndrome were excluded because of a compliance of <33%. The compliance of the participants with 22q11.2 deletion syndrome in the analysis sample was 95% (healthy controls: 98%).	Not reported
Gosens et al. (2023)	To evaluate the effectiveness of a treatment program for substance abuse in individuals with mild intellectual disabilities or borderline intellectual functioning and substance use disorder	Average compliance was 71% (range: 33%-97%). Participants with more no-shows to treatment sessions or periods without treatment had a lower compliance rate.	Not reported

Table 2. continued.

Authors	Study aim(s)	Compliance and drop-out	Acceptability and feasibility of ESM
Hulsmans et al. (2023)	To explore the feasibility of an experience sampling application, with both standardised and personalised items, for adolescents and young adults with a mild intellectual disability or borderline intellectual functioning who receive care in either an outpatient, residential or juvenile detention setting.	Average compliance was 70%. Thirteen participants (26%) dropped out of the study. Participants who completed the 60-day study period had a daily compliance of 86%. Compliance was significantly higher (p <.001) among participants who used their own mobile phone, compared to participants who used a group device for completing the diaries. Most prominent reasons for non-compliance (among the participants who did not drop-out) included: forgetting to fill in the diaries (23%), being in the company of others (20%), and experiencing stress (20%). Things that would have helped participants to increase compliance included: more reminders by care professionals (23%), more app reminders (prompts) (11%), and questionnaire items being updated or altered during the study (14%). Reasons for dropping out of the study were: forgetting to fill in the diary (n = 4), frustrations with technical issues with the device or app (n = 3), stress in their personal lives unrelated to the study (n = 3), and the study being too intense (n = 1) or too boring (n = 1).	Most participants were positive about their participation in the study: 39% of participants experienced it as pleasant, 22% as very pleasant, 37% as neutral, and 2% as unpleasant. The majority of participants (94%) would recommend their peers to participate. Sixty-nine percent of participants indicated that the study duration (60 days) was appropriate, 14% thought this was too short, and for 16% of participants, this was too long. For most participants, one prompt per day was appropriate (73%), 18% indicated that more prompts would have been better, and 8% preferred less than one prompt per day. Perceived gains from participating were: increased self-awareness (64%), receiving a gift-card (30%), reducing problematic behaviours (18%), and better communication with caregivers (18%). Ten percent of participants reported not having gained anything. Perceived drawbacks from participating were: technical issues with the device or app (18%), irrelevance of questionnaire items (14%), or difficulties finding an appropriate moment for completing the diary (8%). More than half (52%) of participants did not report any drawbacks.
Hyde et al. (2021)	To investigate the within-person day-to-day associations between subjective sleep and affect in adults with 22q11.2 deletion syndrome.	Two participants (7%) with 22q11.2 deletion syndrome did not fill in the questionnaires about sleep and were therefore excluded. Average compliance not reported.	Not reported

Table 2. continued.

Authors	Study aim(s)	Compliance and drop-out	Acceptability and feasibility of ESM
Price et al. (2015)	To investigate how cognitive fatigue can be accurately assessed in real-time through the use of mobile assistive technology.	Not reported	Not reported
Schneider et al. (2020)	To investigate affect, stress, and affective reactivity to daily environmental stressors in adults with 22q11.2 deletion syndrome.	Four participants (13%) with 22q11.2 deletion syndrome were excluded because of a compliance of <33%. The compliance of the participants in the analysis sample was 68% (healthy controls: 78%).	Researchers' notes: a subgroup of participants required additional monitoring.
Wilson et al. (2020)	To explore feasibility, reliability, and validity of an experience sampling application for people with intellectual disability.	The average compliance was 34%. The compliance of participants who lived independently was higher compared to those who lived with family (35% vs. 22%, $U = 55$, $p < 0.05$). The compliance of participants who were unemployed was higher than those who were employed (69% vs. 25%, $U = 3$, $p < 0.05$).	Participants enjoyed taking part in the study and reported willingness to participate in similar studies. Most participants found the device easy to use. Reasons for non-compliance included: being elsewhere, not hearing the notifications, or technical difficulties with the device. Participants and parents of participants recommended longer and louder prompts, specific language, and pictographs to aid understanding. Adequate internal consistency and face validity of the experience sampling data.

Table 3. Overview of Themes and Subthemes.

Theme	Subtheme	Example codes
Acceptability	Value and usefulness	I am positive about using the app - it makes me aware of my emotions
		I do not need the app - I can talk about my feelings
	Practical feasibility	I prefer notifications at set times, rather than at random
		I did not enjoy using the app during work hours
Availability	Installation and registration	I received help with installing the app
		The app was easy to find in the app store
	Language and text	The questions were comprehensible
		There was too much text
	Response scales	I found it difficult to understand the response options (Likert scale)
		It was easy to give a grade to my current feelings (visual-analogue scale)
Appropriateness	Adapting the experience sampling design to the needs of users	I did not like getting the same questions every time
		The amount of questions should be adapted to the user's needs
	Adapting the questionnaire response scales to the needs of users	The use of emoticons would distract me from the questionnaire
		Pictograms or emoticons should be used for people with literacy problems

Note. The main themes were defined prior to analysis and were based on the Access to Care Framework.[23]

Acceptability

The theme of acceptability describes the extent to which participants were satisfied with the application. Sub-themes were: value and usefulness, and practical feasibility.

Value and usefulness. This sub-theme describes the general, positive or negative, value participants ascribed to using experience sampling, and the extent to which participants found it useful. Participants did not directly mention any differences in satisfaction between the two apps that they tested. Four participants were generally satisfied with using experience sampling. These participants mentioned that using the application made them aware of their current emotions. As one participant mentioned:

"I am not usually aware of what I am feeling. Using the app made me realise that I was not feeling great." [P4]

This participant also recognised the value of experience sampling, compared to retrospective questions:

"'How are you feeling right now?' I think that was great. Sometimes, when you get a question about how you were feeling, you really have to think hard: 'How was I feeling?' These questions were about exactly that moment. That is important to ask. Not: 'How were you feeling ten minutes ago?'" [P4]

Three participants mentioned that experience sampling may be useful for themselves and other people who cannot easily talk about their feelings. Relatedly, two participants mentioned that it would be useful for clients to share the questionnaire results with direct support professionals, to provide insight into their current states. One participant did this during the study period and found it useful:

"Sometimes, I am not able to tell her exactly how I feel. Using the app would be, you know, a good alternative." [P3]

Three participants were not satisfied with the experience sampling method. These participants found it difficult to respond to the questionnaire items, because of problems with describing their feelings. One participant, who during the interviews mentioned being autistic, found the questionnaire items too abstract:

"I find it very hard to assess how I am feeling, or how I am doing, or how I should respond to the questions". [P6]

Lastly, two participants did not see value in using the app for themselves, but would be willing to use it again in the context of research, thus, to help other people. As one participant reported:

"Look, I would not use it for myself. But it was great to test it for other people. "

Practical feasibility. This sub-theme describes the feasibility of using an experience sampling application in day-to-day life, including the amount of notifications per day, the schedule of prompts during the day (set times or at random), and the length of the questionnaires. In this study, participants were prompted four to five times per day to fill in the questionnaire. Participants agreed that more than five times would be too much of a burden. Two participants mentioned that three times per day would be enough, as illustrated by the following quotes:

> *"More than five times? I would think: 'Oh, here is the app again'… I do not want to spend my entire day responding to questions about my feelings." [P5]*

> *"To be fair, I found it too much. It already started in the morning, around 6:30am. That was really early. I thought to myself: 'I am not going to do that'. At around 9am, I was like, 'I am awake now, I can do this now.'" [P7]*

With regard to the time schedule of prompts, four participants strongly preferred receiving prompts at set times, rather than at random times. Three participants did not mention a preference for random versus scheduled prompts. One participant reported having problems with attention, and found the unpredictability of the random prompts difficult:

> *"I have ADHD [attention-deficit hyperactivity disorder]. And I found it very difficult to receive prompts at the most random and unwanted times. […] For me, scheduled prompts would be very important." [P4]*

Participants seemed satisfied with the length of the questionnaires – all reported having finished the questionnaires within approximately five minutes. Most of the participants reported having completed the majority of the questionnaires. Reasons why they sometimes were not able to respond included being at work, doing something else, or having too much on their mind. One participant found it difficult to switch between tasks when being at work:

"When I am at work, I am doing other things. I just do not like filling in a questionnaire when I am in between tasks. Having two things to focus on is difficult for me, you know." [P2]

Another participant felt too stressed one evening to respond to the questionnaires, and ignored the prompts:

"Yesterday, there were way too many… I was like, screw this. I was too distressed." [P6]

Availability

The theme of availability describes the capacity of participants to respond to the questionnaire items. Sub-themes were: installation and registration, language and text, and response scales.

Installation and registration. This sub-theme describes the ease of installing and registering the applications. Four participants mentioned having received help with installation and registration. According to the participants, the applications were easy to find in their smartphone's app store. Other than Ethica and m-Path, the Quenza application asks for a password before the start of each questionnaire, which was a barrier for one participant:

"The first time I was afraid I used the wrong password. But it ended up being the correct one." [P2]

Language and text. This sub-theme describes participants' views on the language and text used in the questionnaire. Five participants found the language comprehensible and straightforward. Two participants indicated that there was too much text in the questionnaire, referring to the instruction texts and the questionnaires with Likert response scales:

"There were too many answers to choose from. But this may have to do with my preference for the numbers [on the visual-analogue scale]." [P3]

The items about positive and negative affect were clear to most participants. One participant indicated that the positive affect items were too similar:

> *"Many people may find it difficult to distinguish 'joyful' from 'cheerful', these are too similar. This was not a problem for me, but it may be for others." [P4]*

Response scales. This sub-theme refers to participants' experiences with the Likert and visual-analogue response scales. Four participants preferred the visual-analogue scales over the Likert scales, because they did not have to think about the meaning of response items. As one participant mentioned:

> *"It is better with the numbers, because you can just use a slider [from 0 to 10]. You do not have to think about the meaning of 'a little bit' or 'very much'". [P3]*

The Likert response scale included the word 'neutral', which was difficult to understand for two participants. Their direct support professional explained the word to them, after which they were able to continue with the questionnaire. One participant preferred the Likert response scale over the visual-analogue scale, because she found the items clearer:

> *"Yes, these answers were much more clear to me. This was easier for me, because I did not have to think about: 'What is a 6?' [on the visual-analogue scale]. This was much easier for me." [P4]*

Two participants did not like to respond to the questions about social contact, which were scheduled after the questions about affect. For these participants, these questions felt too much of a burden. Participants were divided about the use of open text fields in the questions about social contact. For one participant, typing was difficult. The other participants did not mention having difficulties with typing. One participant expressed the wish for including follow-up open questions about affect:

"I found the tests a bit monotonous, because the questions were the same every time. The only possible answers were: 'I'm feeling good' or 'I'm not feeling good'. There were no options for describing why you were feeling a certain way." [P5]

Appropriateness

The theme of appropriateness refers to ways in which the experience sampling design and the questionnaire response scales could be adapted to the needs of different users.

Adapting the experience sampling design to the needs of users. This sub-theme describes participants' views on how the experience sampling design could be adapted to the needs of users. Two participants mentioned that the use of rewards would increase motivation, such as receiving points or prizes after completing questionnaires. But for one participant, this may also lead to disappointment:

"It depends… Sometimes, this would be helpful and motivating, but if I am not able to finish the questionnaire, it would make me feel worse." [P3]

One participant found it annoying to respond to the same questions every time:

"And those questions about… Are you with someone right now? Those questions were the same every time. That was annoying." [P2]

This participant also mentioned that the number of questions should be adapted to the capacity of the user:

"I would not add any more questions. […] Well, the questions are fine as they are, for me, but for someone else, this might be too much, and then it becomes too difficult." [P2]

Adapting the questionnaire response scales to the needs of users. This sub-theme describes participants' views on how the response scales of questionnaires could be adapted to the needs of users, including the use of emoticons and other visuals. Participants were divided about the use of emoticons to reflect positive and negative affect. Emoticons were not included in the current questionnaires but participants were asked if this would aid understanding of the questions. Four participants indicated that it would be useful for them, or others, to support the questions with emoticons. For one participant, this would be distracting:

> *"Emoticons would distract me too much from the questions. […] That would be too complicated for me, to read the questions and also try to understand the meaning of the emoticons." [P2]*

Another participant did prefer the use of pictograms over written text:

> *"Picto's are better for me. Even though I can read and write well, I am intelligent, and I have studied […]. But visual material just works better for me." [P6]*

Participants agreed that the preference or need for including emoticons would differ between people. Three participants indicated that it would be important to include visuals for people with literacy problems. As one participant indicated:

> *"For some clients, emoticons may be useful. And for clients who cannot read, pictograms could be used. But for me, I can read, so emoticons or pictograms are not necessary." [P1]*

One participant mentioned that experience sampling applications should be inclusive to both people with and without literacy problems:

> *"If you would design one app… Well… It depends on the user. If this is someone who cannot read, then they should be able to adjust the settings and use pictograms instead of written text." [P2]*

Discussion

The scoping review identified seven studies in which experience sampling was used for assessing mental health of people with intellectual disability, all of which were published from 2020 onwards. Compliance to the questionnaires was reported by three studies, and ranged between 34% to 95%.[3,22] Two studies were specifically aimed to examine the acceptability and feasibility of experience sampling.[3,25] In both studies, participants used an experience sampling application to report on mental health related outcomes, for 7 consecutive days[3] and 60 consecutive days,[25] and reported on their experiences during follow-up interviews. Wilson and colleagues (2020) mostly described participants' experiences with practical and technical difficulties, such as the comprehensibility of the questions and reasons for non-compliance. Hulsmans et al. (2023) provided a more comprehensive description of participants' experiences and views regarding the study duration, number of prompts, and perceived gains and drawbacks from participating, in addition to reasons for non-compliance and drop-out. Thus, findings regarding the acceptability and feasibility of experience sampling for this population have just started to emerge. Because the field of intensive longitudinal research has been rapidly evolving in the past decades,[7] more experience sampling studies including people with intellectual disability may be expected.

Interviews with adults with mild intellectual disability using experience sampling applications revealed a more rich and differentiated set of experiences. Within the first of the three domains from the Access to Care Framework,[15] acceptability, sub-themes were value and usefulness and practical feasibility. Participants were divided about the usefulness of experience sampling for their own ability to communicate about feelings, and reported on reasons why using an experience sampling application would not fit well into their daily lives. For the domain availability, sub-themes were installation and registration, language and text, and response scales. The applications were mostly easy to find and install. Participants expressed different preferences for the use of visual-analogue or Likert response scales. For the domain appropriateness, sub-themes were adapting the experience sampling design to the needs of users and adapting

the questionnaire response scales to the needs of users. Some participants found it annoying to receive random prompts, for example, because of problems with attention, and some valued the predictability of the same order of questions during every prompt, whereas others found this boring and annoying. Further, we explored the acceptability of using pictograms or emoticons in the questionnaires. Some participants indicated that the use of visuals depends on the needs and preferences of individuals.

In the scoping review, the compliance rates varied widely between studies. The lowest compliance rate of 34% was reported by Wilson et al. (2020). Other studies reported higher compliance rates, around 70%.[24–26] Feller et al. (2021) reported a relatively high response rate of 95%. As suggested by these authors, this may be due to the guidance the researchers provided during the data collection. A recent meta-analysis showed an average compliance of 79% across k = 347 experience sampling studies involving non-clinical and clinical samples and different age groups.[28] Most of the studies in our review reported comparable response rates, with the exception of Wilson et al. (2020). As the authors noted, this may be due to the inability to use their device or respond to notifications during daily activities, such as work. Participants in our study also mentioned this as a reason for not responding to the questionnaires. However, we did not examine participants' compliance rates, which is a limitation.

Both the scoping review and the interviews reflect the heterogeneity of the target population. Two studies in the scoping were targeted at people with mild intellectual disability and borderline intellectual functioning,[25,26] in one study, people with mild-to-moderate intellectual disability were the target group,[3] and in four studies, the target group were people with 22q11.2 deletion syndrome.[21–24] In the feasibility study by Hulsmans et al. (2023), one-third of participants had at least one comorbid DSM-5 diagnosis. Two participants in our study mentioned having a neurodevelopmental disorder (attention-deficit hyperactive disorder and autism), which impacted their acceptance of the method itself and the experience sampling designs they tested. On the one hand, these findings point to the need to tailor experience sampling methods to the needs and preferences of users. In terms of the domains of accessibility of the Access to Care Framework,[15] this may

make experience sampling methods more acceptable, available, and appropriate for people with intellectual disability. On the other hand, more research may be needed to examine the feasibility of experience sampling methods for people with intellectual disabilities and different comorbid neurodevelopmental conditions. Additionally, none of the studies in the review included individuals with moderate to severe intellectual disability, which is another avenue for future research.

Recommendations for research and clinical practice

We recommend future experience sampling studies involving people with intellectual disability to personalise the study design, following the needs and preferences of individual participants. On the level of questionnaire items, researchers may consult with participants to assess if they understand the language, which response scales they prefer, and if they prefer the use of visuals, such as pictograms. A possible way to tailor the content of the questionnaire to participants' needs could be to include both standardised and personalised questions, which was previously done by Hulsmans and colleagues (2023). The experience sampling protocol may be adjusted to participants' needs and preferences for the duration of sampling (i.e., the minimum number of days participants find acceptable), number of prompts per day, and their preference for time-scheduled or random prompts. In our study, some participants strongly preferred prompts on a set time schedule because of problems with attention or with having difficulty filling in the questionnaires during other tasks. While using a set time protocol for these participants may increase their acceptance of the method and possibly their compliance, this may also limit possibilities for exploring time-variant associations between internal states and environmental exposures. Furthermore, the way in which set times are chosen may introduce selection bias in the experiences and events that are sampled. In addition, researchers are advised to check in with participants during the study period, to provide support and help with any technical difficulties that may arise. Lasty, we recommend researchers of future experience sampling studies to include a qualitative examination of participants' experiences at the end of the study period. Some participants in our study reported being irritated by the questionnaire items or prompts. This may lead to artefactual reports of negative mood in the questionnaire, causing potential bias in the

study findings. For example, a participant may report feeling annoyed in response to a prompt because of the prompt itself, and not because they had a negative mood due to another situation. Asking participants about their experiences with participating in the study, such as by debriefing interviews, may give some insight in the reliability of the study findings and explain potentially low study compliance.

According to the interviews, participants recognised the value of using experience sampling to keep support staff informed about their mood. Experience sampling might be used for monitoring and informing care, as it provides a quantitative evaluation of mental health from the client perspective. Until now, assessments of mental health using clinical rating scales or parent-reported questionnaires omitted patient perspectives about issues of relevance to their mental health. Additionally, it has been pointed out that the perception of individuals' mental health by clinicians and clients themselves differ.[29,30] The validity of proxy reflections of unobservable internal states is limited, as the personal perspective can only truly be understood by the individual's self report.[31] The use of experience sampling methods might thus be a user-friendly way to acquire information from the people with intellectual disabilities themselves, which is also encouraged by regulatory authorities such as the Food and Drug Administration and European Medicines Agency.[32,33] While the interviews pointed towards similar applications of experience sampling as a source of practice-based evidence, the scoping review did not turn up research in which experience sampling was evaluated on potential benefits for diagnosis and treatment. However, experience sampling methods can be used as an outcome measurement instrument for interventional studies to relate therapeutic findings to the self-evaluated mental functioning. Such research may build on the extensive tradition in the field of single-case experimental designs, which have served the dual purpose of testing mechanisms of clinical change and clinical intervention.[34] Due to the heterogeneity of the intellectual disability population, multiple data points are recommended in these types of studies to increase the study's validity and to be able to observe fluctuations over time,[34–36] which is enabled by experience sampling methods.

Strengths and limitations

A strength of the current study was the involvement of experts-by-experience in both the design and participation in the interview study. These individuals are trained to be involved in various stages of research projects and are able to represent a broader group of clients with intellectual disability receiving care. Another strength was the use of three different experience sampling applications and two different response scales (visual-analogue and Likert scales), which allowed participants to experience and report on differences in design. However, none of the participants mentioned any differences in design between the applications, other than their preference for either the visual-analogue or the Likert response scale.

In the scoping review, we did not find studies with people with severe or profound intellectual disability, although these were not excluded based on the criteria. All of the included studies reported text-based experience sampling questionnaires, indicating that these were not suitable for people with severe or profound intellectual disability, and/or that this group was excluded from participation. Relatedly, all interview participants were adults with a mild intellectual disability. Therefore, the findings are not generalisable to adolescents or elderly, and people with moderate, severe, or profound intellectual disability. Another limitation was that we did not specifically ask the interview participants to consider the use of experience sampling for research or clinical purposes. This would have led to more differentiated responses and specific recommendations for tailoring experience sampling methods for research and clinical practice. Lastly, our study included relatively few measurements: around 24-30 prompts per individual depending on the application, compared to 60[25] and 48 measurements[3] in other feasibility studies.

Conclusions

This study showed important gaps in research on the accessibility of experience sampling methods for mental health of people with intellectual disability. Based on a scoping review and interviews with stakeholders, we recommend future studies to tailor experience sampling designs to the

needs and preferences of individual participants, both on the level of the questionnaire items (wording, response scales, use of visuals) and in the experience sampling protocol (duration of sampling, number of prompts per day, set-time schedule vs. random prompts). In addition, researchers may include a qualitative examination of participants' experiences with using this method (e.g., debriefing interviews), to reveal potential bias in the study findings. In addition to mental health research, experience sampling may be used in clinical practice to contribute to personalised care, although more work is needed in this area.

Acknowledgements

We thank Benjamin van Ark for his consultation in the design of the experience sampling protocol.

Funding

This study was funded by a Mental Health Alliance Fund from Amsterdam Public Health research institute (reference number: 2021136/MdB/EdB).

Conflicts of interest

The authors have no conflicts of interest to declare.

Authors' contributions

LB was involved in the study concept and design, had a major role in acquisition and interpretation of data, and drafted the manuscript. CP and AM had a major role in the acquisition of data and drafted and revised the manuscript. AvE revised the manuscript for content. CS was involved in the study concept and design, and drafted and revised the manuscript.

References

1. Kooijmans R, Mercera G, Langdon PE, Moonen X. The Adaptation of Self-Report Measures to the Needs of People With Intellectual Disabilities: A Systematic Review. Clin Psychol Sci Pract. 2022;29(3).
2. Hughes-McCormack LA, Rydzewska E, Henderson A, MacIntyre C, Rintoul J, Cooper SA. Prevalence of mental health conditions and relationship with general health in a whole-country population of people with intellectual disabilities compared with the general population. BJPsych Open. 2017;3(5).
3. Wilson NJ, Chen YW, Mahoney N, Buchanan A, Marks A, Cordier R. Experience sampling method and the everyday experiences of adults with intellectual disability: A feasibility study. J Appl Res Intellect Disabil. 2020;33(6).
4. Hektner J, Schmidt J, Csikszentmihalyi M. Experience Sampling Method. SAGE Publications, Inc.; 2007.
5. de Vries LP, Baselmans BML, Bartels M. Smartphone-Based Ecological Momentary Assessment of Well-Being: A Systematic Review and Recommendations for Future Studies. Vol. 22, Journal of Happiness Studies. 2021.
6. Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, Van Os J. Experience sampling research in psychopathology: Opening the black box of daily life. Psychol Med. 2009;39(9).
7. Schick A, Rauschenberg C, Ader L, Daemen M, Wieland LM, Paetzold I, et al. Novel digital methods for gathering intensive time series data in mental health research: Scoping review of a rapidly evolving field. Vol. 53, Psychological Medicine. 2023.
8. Boemo T, Nieto I, Vazquez C, Sanchez-Lopez A. Relations between emotion regulation strategies and affect in daily life: A systematic review and meta-analysis of studies using ecological momentary assessments. Vol. 139, Neuroscience and Biobehavioral Reviews. 2022.
9. Hall M, Scherner P V., Kreidel Y, Rubel JA. A Systematic Review of Momentary Assessment Designs for Mood and Anxiety Symptoms. Vol. 12, Frontiers in Psychology. 2021.
10. Shiffman S, Stone AA, Hufford MR. Ecological Momentary Assessment. Annu Rev Clin Psychol. 2008;4(1):1–32.
11. Fujiura GT. Self-reported health of people with intellectual disability. Intellect Dev Disabil. 2012;50(4).
12. Hoffmann F, Banzhaf C, Kanske P, Bermpohl F, Singer T. Where the depressed mind wanders: Self-generated thought patterns as assessed through experience sampling as a state marker of depression. J Affect Disord. 2016;198.
13. Gloster AT, Miché M, Wersebe H, Mikoteit T, Hoyer J, Imboden C, et al. Daily fluctuation of emotions and memories thereof: Design and methods of an experience sampling study of major depression, social phobia, and controls. Int J Methods Psychiatr Res. 2017;26(3).
14. Daros AR, Daniel KE, Meyer MJ, Chow PI, Barnes LE, Teachman BA. Impact of social anxiety and social context on college students' emotion regulation strategy use: An experience sampling study. Motiv Emot. 2019;43(5).
15. Levesque JF, Harris MF, Russell G. Patient-centred access to health care: Conceptualising access at the interface of health systems and populations. Int J Equity Health. 2013;12(1).
16. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015: Elaboration and explanation. Vol. 349, BMJ (Online). 2015.
17. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016.
18. Ebesutani C, Regan J, Smith A, Reise S, Higa-McMillan C, Chorpita BF. The 10-item Positive and Negative Affect Schedule for Children, Child and parent shortened versions: Application of item response theory for more efficient assessment. J Psychopathol Behav Assess. 2012;34(2).

19. Peeters FPML, Ponds RWHM, Vermeeren MTG. AFFECTIVITEIT EN ZELFBEOORDELING VAN DEPRESSIE EN ANGST. Tijdschr Psychiatr. 1996;38(3).
20. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol. 2006;3(2).
21. Feller C, Ilen L, Eliez S, Schneider M. Characterizing Daily-Life Social Interactions in Adolescents and Young Adults with Neurodevelopmental Disorders: A Comparison Between Individuals with Autism Spectrum Disorders and 22q11.2 Deletion Syndrome. J Autism Dev Disord. 2023;53(1).
22. Feller C, Ilen L, Eliez S, Schneider M. Psychotic experiences in daily-life in adolescents and young adults with 22q11.2 deletion syndrome: An Ecological Momentary Assessment study. Schizophr Res. 2021;238.
23. Hyde J, Eidels A, van Amelsvoort T, Myin-Germeys I, Campbell L. Gene Deletion and Sleep Depletion: Exploring the Relationship Between Sleep and Affect in 22q11.2 Deletion Syndrome. J Genet Psychol. 2021;182(5).
24. Schneider M, Vaessen T, van Duin EDA, Kasanova Z, Viechtbauer W, Reininghaus U, et al. Affective and psychotic reactivity to daily-life stress in adults with 22q11DS: a study using the experience sampling method. J Neurodev Disord. 2020;12(1).
25. Hulsmans DHG, Poelen EAP, Lichtwarck-Aschoff A, Otten R. The feasibility of daily monitoring in adolescents and young adults with mild intellectual disability or borderline intellectual functioning. J Appl Res Intellect Disabil. 2023;36(4).
26. Gosens LC, Poelen EA, Didden R, de Jonge JM, Schellekens AF, VanDerNagel JE, et al. Evaluating the effectiveness of Take it Personal!+ in people with mild intellectual disability or borderline intellectual functioning and substance use disorder: A multiple baseline single-case experimental study. Behav Ther. 2023;In press.
27. Price E, Moore G, Galway L, Linden M. Towards a mobile assistive technology for monitoring and assessing cognitive fatigue in individuals with acquired brain injury. In: Proceedings - 15th IEEE International Conference on Computer and Information Technology, CIT 2015, 14th IEEE International Conference on Ubiquitous Computing and Communications, IUCC 2015, 13th IEEE International Conference on Dependable, Autonomic and Se. 2015.
28. Wrzus C, Neubauer AB. Ecological Momentary Assessment: A Meta-Analysis on Designs, Samples, and Compliance Across Research Fields. Assessment. 2023;30(3).
29. Janse AJ, Gemke RJBJ, Uiterwaal CSPM, Van Der Tweel I, Kimpen JLL, Sinnema G. Quality of life: Patients and doctors don't always agree: A meta-analysis. J Clin Epidemiol. 2004;57(7).
30. Lunsky Y, Bramston P. A preliminary study of perceived stress in adults with intellectual disabilities according to self-report and informant ratings. J Intellect Dev Disabil. 2006;31(1).
31. Maguire S, Davison J, McLaughlin M, Simms V, Bunting B. Exploring the psychometric properties of self-report instruments used to measure health-related quality of life and subjective wellbeing of adolescents with intellectual disabilities: A Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) systematic review. J Appl Res Intellect Disabil [Internet]. 2023 Apr 26; Available from: http://www.ncbi.nlm.nih.gov/pubmed/37101341.
32. Matts ST, Webber CM, Bocell FD, Caldwell B, Chen AL, Tarver ME. Inclusion of patient-reported outcome instruments in US FDA medical device marketing authorizations. J Patient-Reported Outcomes. 2022;6(1).
33. US Food and Drug Administration. Guidance for Industry Patient-Reported Outcome Measures Use in Medical Product Development to Support Labeling Claims. Clinical/Medical Federal Register. 2009.
34. Müller AR, Brands MM, van de Ven PM, Roes KC, Cornel MC, van Karnebeek CD, et al. The Power of 1: Systematic Review of N-of-1 Studies in Rare Genetic Neurodevelopmental Disorders. Neurology. 2021.
35. Backman CL, Harris SR. Case studies, single-subject research, and N of 1 randomized trials: Comparisons and contrasts. Am J Phys Med Rehabil. 1999.
36. Guyatt G, Sackett D, Adachi J, Roberts R, Chong J, Rosenbloom D, et al. A clinician's guide for conducting randomized trials in individual patients. CMAJ. 1988.

PART III

Genetic diagnostics and implementation in intellectual disability care

Chapter 8

Do we care? Reporting of genetic diagnoses in multidisciplinary intellectual disability care: a retrospective chart review

Annelieke R. Müller, Erik Boot, Stijn B. Notermans, Carlo Schuengel, Lidewij Henneman, Martina C. Cornel, Mieke M. van Haelst, Mariëlle Alders, Clara D.M. van Karnebeek, Bas Bijl, Frits A. Wijburg, Agnies M. van Eeghen

Submitted

Abstract

Purpose: Advances in understanding the etiology of intellectual disability (ID) has led to insights in potential (targeted) treatments and personalized care. Implications of ID on health are often complex and require a multidisciplinary approach. The aim was to investigate the reporting of genetic diagnoses in multidisciplinary ID care and to identify associated clinical and demographic factors.

Methods: A retrospective chart review was performed on a randomly selected sample of clients (n=380) of a large ID care organization in the Netherlands. Data on genetic etiology, including genetic testing and diagnoses, and clinical and demographic characteristics were collected from files held by multidisciplinary team members.

Results: Reports on genetic etiology were available in 40% of the study sample (n=151), with a genetic diagnosis recorded in 34% (n=51). In those with reported genetic diagnoses, this was reported in 90% of medical, 39% of psychodiagnostic, and 75% of professional caregivers' files. Older age, mild ID, and the legal representative not being a family member were associated with less reported information on genetic etiology.

Conclusion: This study revealed that genetic diagnoses were often not reported in ID care files. Recommendations were formulated to reduce delay in diagnosis, and improve personalized care for individuals with ID.

Introduction

About 1-3% of the population is affected with intellectual disability (ID),[1] which is characterized by substantial limitations in both intellectual functioning and adaptive behavior, originating during the developmental period.[2] Due to rapid technological advances, a genetic diagnosis can be identified in up to 50% of individuals with ID, although estimates of the diagnostic yield vary considerably across studies.[3,4] Currently, more than 1,500 monogenic causes of ID are known in addition to other causes like copy number variations (CNVs).[5] Such genetic neurodevelopmental disorders and neurometabolic disorders often manifest with complex and variable multiorgan comorbidity. As a result, many different healthcare providers (HCPs) are usually involved in multidisciplinary care, including physicians, psychologists, and professional caregivers.

Knowing the cause of ID provides information about associated somatic and neuropsychiatric manifestations and may lead to targets for prognosis, screening, prevention, monitoring and treatment.[6] Moreover, it may result in increased life expectancy for those affected. Together with improved genetic diagnostics, targeted treatments and disorder-specific care are increasingly available,[7,8] allowing for personalized care, which is the implementation of etiology-driven health monitoring and treatments.[9] Disorder-specific care is illustrated by anticipatory care planning for individuals with Down syndrome who eventually all show neuropathological changes of Alzheimer's disease by the age of forty.[10] Research has mainly focused on pediatric ID, although a diagnosis may provide benefits for adults too. More knowledge of complex neuropsychiatric manifestations, the greatest burden of most rare genetic neurodevelopmental disorders,[11,12] can improve targeted neuropsychological examination, psychoeducation, and behavioral interventions.[13,14] Disorder-specific guidelines are increasingly available, providing recommendations for medical, social, psychiatric and behavioral care.[15]

Although a genetic diagnosis may thus provide important benefits for affected individuals and their families, it is unknown to what extent genetic diagnoses, including information on phenotype and management, are integrated into multidisciplinary ID care.

We therefore evaluated the integration of personalized care across disciplines in ID care to identify care gaps and targets for improvements. The primary objective was to investigate the reporting of information on genetic etiology, including both genetic testing and genetic diagnoses in medical files, files used by psychologists and behavioral experts and therapists, and files used by professional caregivers. A secondary objective was to examine how often, at what time and how detailed, information on the genetic diagnosis was available in these files, and to investigate associated clinical and demographic factors to assess the integration of genetic diagnoses into clinical multidisciplinary care.

Methods

Study design and setting

This was a retrospective chart review at 's Heeren Loo, the largest care organization for ID in the Netherlands. We systematically recorded data on genetic etiology, including genetic testing and diagnosis, and clinical and demographic characteristics collected from files held by multidisciplinary team members involved.

Care systems and study population

Individuals receiving support or care from 's Heeren Loo are registered in the electronic care system (Figure 1). The system is used by all involved HCPs including behavioral scientists, to report on paramedical care, supported living, and other support. It also includes information about legal representatives. For medical care by general practitioners and ID physicians, another electronic care system is widely used.

Individuals who were registered in the electronic care system and received at least two months of support or care from 's Heeren Loo prior to data extraction were included in the study, unless there was no consent for using their data for research. Individuals who visited an expertise center genetic syndromes at 's Heeren Loo were also excluded from analyses, as a genetic diagnosis is a prerequisite for receiving care by this center.

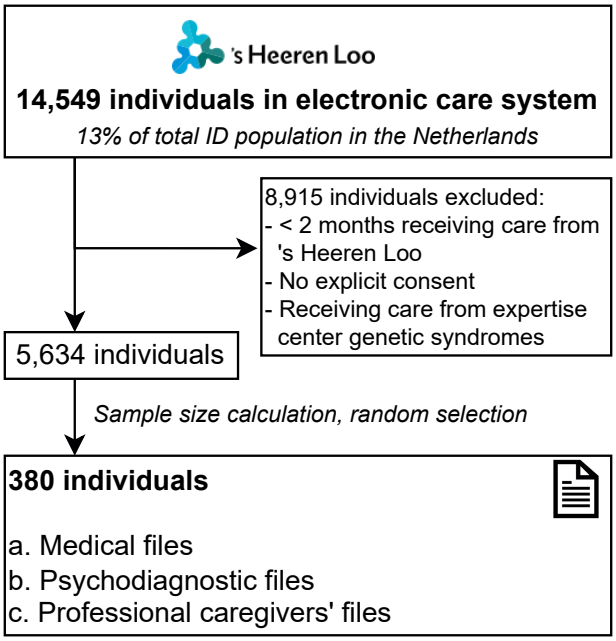


Figure 1. Flow diagram depicting the selection procedure of individuals eligible for electronic care file search.

Procedures

The electronic care system of 's Heeren Loo contained data on 14,549 individuals (accessed at September 6, 2021), which is approximately 13% of the total ID population in the Netherlands.[16] Of these, around 6,200 individuals live within sheltered care facilities of the organization. The sample size for this study was based on a population size of 14,549 individuals with a standard deviation of 50%, a sampling error of 5%, and confidence interval of 95%, yielding a required sample size of ≥374 individuals. We included 380 individuals who met our inclusion criteria.

Electronic care files were searched for the following demographic information including type of support and care: age, sex, whether the individual had a legal representative and their relationship to the individual (e.g., family member or non-family member such as professional or friends), ID physician involvement, living situation and whether there was medical care on site. Files were reviewed to assess what information was provided on the following clinical information: severity of ID based on clinical

assessments, intelligence tests and adaptive behavior assessments,[2] genetic diagnosis, whether genetic diagnostics were performed, details on genetic test results if applicable, and at what age, referred by whom, total amount of genetic tests (until diagnosis), type of etiological (genetic or metabolic) test performed, the year in which it was performed, test results including classification of pathogenicity and genetic diagnosis causing the ID. A diagnosis was considered confirmed if a genetic test was performed and the specific diagnosis was confirmed with a letter from a clinical geneticist or with an available genetic test result. If no such report was available, it was assessed whether the genetic diagnosis was likely, based on available information and expert opinion (MvH, MA, AM, SN, AvE). Furthermore, given that separate files are used by specific HCPs, it was noted whether 1) medical files, 2) psychodiagnostic files used by psychologists, behavioral experts and therapists, and/or 3) files used by professional caregivers including daily care records and individual support plans were available and whether the information on genetic etiology was mentioned in these files (Figure 1). Also, it was noted whether the involved physician was ID physician or general practitioner.

Data analyses

Demographic and clinical characteristics were described and it was examined to what extent information on genetic etiology was available in medical files, psychodiagnostic files, and files used by professional caregivers. If a genetic cause for ID was reported in any file, it was examined whether this information was also available in files from the other disciplines as well. Independent t-tests, analyses of variance (ANOVAs), and chi-squared tests were performed to investigate whether demographic and clinical variables were associated with availability of information on genetic etiology. These variables included medical care on site, ID physician part of care team, living situation, legal representative, age, sex, and level of ID. Mann-Whitney U tests, Kruskal-Wallis or Fisher's exact tests were used when assumptions for parametric analyses were not met. Post hoc analysis was performed using cell-wise adjusted standardized residual analysis with a Bonferroni adjusted α. A logistic regression analysis was performed to ascertain the relative effects of the associated variables on the likelihood that individuals have information on genetic etiology reported,

chosen from their statistical significance on bivariate analyses. As for the legal representative variable, those with a family or non-family member as legal representative were included in the model, excluding those without a (reported) legal representative to reduce possible bias or multicollinearity. Cochran's Q test was used to examine to what extent genetic test results were available or recorded in either medical files, psychodiagnostic files or files used by caregivers. Statistical analyses were performed using SPSS version 28.0 (SPSS Inc., Chicago, IL, USA) with a two-sided significance level of 5%.

Ethics Declaration

A waiver for formal approval was obtained from the Institutional Review Board of Amsterdam UMC, the Netherlands (#W21_398#21.443). Individuals receiving care or support from 's Heeren Loo do or do not provide consent for using their data for research. If there was no consent for using their data, individuals were excluded. All data were collected and analyzed according to the EU General Data Protection Regulation and the Dutch General Regulation Data Protection (Uitvoeringswet AVG). The study used existing data that were handled anonymously as much as possible. For example, we used age categories instead of the exact age. The study adheres to the principles of the Declaration of Helsinki.

Results

We included 380 individuals at a median age of 46 (interquartile range 31; range 9-95) years old. Demographic and clinical characteristics are presented in Table 1.

Table 1. Demographic and clinical characteristics of the study sample.

	Total study sample		Information available on genetic etiology	
	N = 380		N = 151	
Demographics	N	%	N	%
Age				
<18 years	19	5.0	9	6.0
≥18 years	361	95.0	142	94.0
Sex, female	180	47.4	70	46.4
Legal representative				
None	45	11.8	10	6.6
Family member	249	65.5	118	78.1
Professional/other	86	22.6	23	15.2
ID physician involved	253	66.6	102	67.5
24h/day care	313	82.3	121	80.1
Clinical				
Severity of ID [a]				
Mild	110	28.9	22	14.8
Moderate	130	34.2	56	37.6
Severe	97	25.5	54	36.2
Profound	41	10.8	17	11.4

[a] Severity of ID was unknown in two individuals.

ID, intellectual disability.

Genetic diagnoses in multidisciplinary ID care

Of the total study sample, information on genetic etiology was reported in the electronic care system of 151 individuals (40%) (Figure 2). If information was recorded, most often it concerned negative test results (64/151), followed by a genetic diagnosis (51/151), variants of uncertain significance (VUS) (21/151), and clinical diagnosis or genetic variants mentioned as cause of ID although insufficiently or incorrectly described to be considered a genetic diagnosis (15/151). Particularly, in psychodiagnostics files or files used by professional caregivers, these VUS or genetic variants not considered a genetic diagnosis were mentioned as cause of the ID. The information on reported genetic cause of ID was reported in the different care files used by physicians (90%), psychologists (39%), and professional caregivers (75%) (Figure 3). Cochran's Q test revealed a statistically significant difference in

the proportion of information reported in the care files on genetic cause of ID across the three types of care files ($X^2(2, N = 87) = 68.698$, $p < .001$), with physicians reporting more often compared to psychologists ($p < .001$), and professional caregivers ($p = .025$), and professional caregivers reporting more compared to psychologists ($p < .001$).

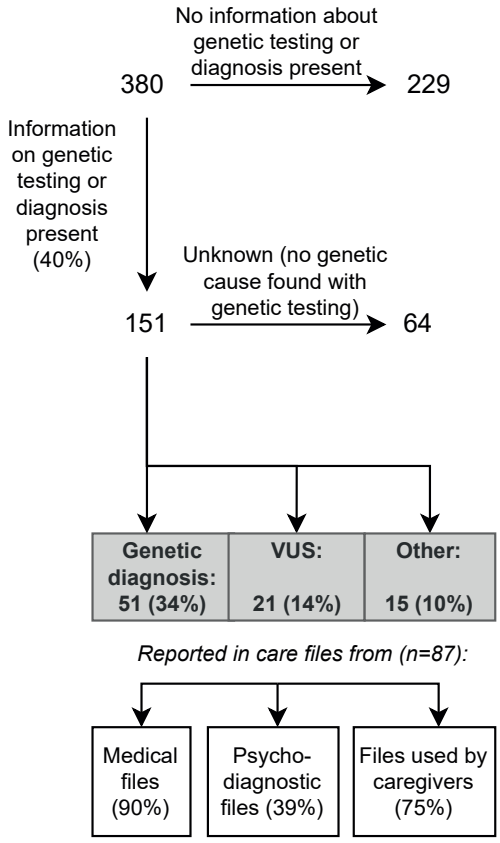


Figure 2. Flow diagram of sample with frequency of information on genetic etiology present, including genetic diagnostics and results, according to care files from different disciplines. 'Other' includes clinical diagnosis or genetic variants mentioned in care files as cause of ID, although insufficiently or incorrectly described to be considered as a genetic diagnosis.

VUS, variant of uncertain significance.

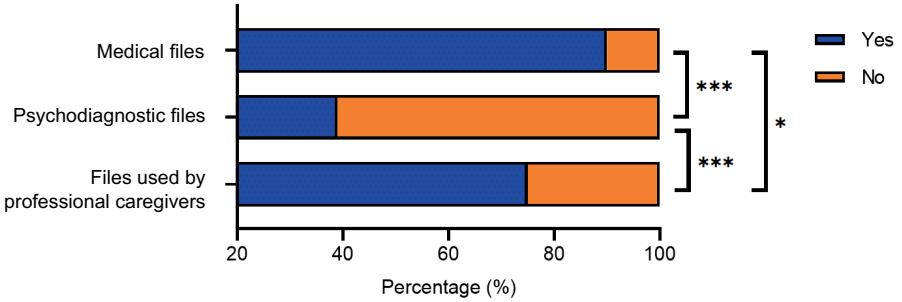


Figure 3. Proportion of genetic diagnoses, variants of uncertain significance, and other (N=87) reported (blue) and missing (orange) in medical files, psychodiagnostic files, or files used by professional caregivers.

* indicates p ≤ .01; *** indicates p ≤ .001.

Factors associated with availability of information on genetic etiology

Significant associations were found between presence of information on genetic etiology and age (r=-0.16, p=.002), the severity of ID (X^2 (3, N = 378) = 28.898, p < .001), and the type of relationship with the legal representative (X^2 (2, N = 380) = 17.323, p < .001) (Supplementary Table A). Post hoc analysis revealed that individuals with moderate and severe ID were more likely to have information in their files reporting on genetic etiology compared to individuals with mild ID (p < .001). Individuals with a family member as a legal representative were more likely to have this information reported compared to those without a legal representative or those with a non-family member as legal representative (p < .001). No significant associations were found between information on genetic etiology and sex, location of receiving care, presence of medical care on site, and 24 hours support.

The logistic regression model predicting the effects of age, level of ID, and legal representative on the likelihood that individuals have information on genetic etiology reported in any file, was statistically significant (X^2(5) = 48.367, p < .001), explaining 18.2% (Nagelkerke R^2) of the variance in presence of information on genetic etiology and correctly classified 69.1% of cases (Table 2). Increasing age was associated with a decreased likelihood of reporting information on genetic etiology (odds ratio [OR] = 0.971, p < .001). Individuals with a moderate, severe or profound ID were respectively 3.36, 6.32, and 4.02 times more likely to have reported information on genetic etiology compared to individuals with a mild ID. Individuals with a

family member as legal representative were 2.17 times more likely to have reported information on genetic etiology compared to those with a legal representative other than a family member.

Table 2. Factors associated with availability of information on genetic etiology in all files using a logistic regression model (N = 335).

		95% CI of B		
Variable	B	Lower	Upper	p
Age	-.029	.958	.985	<.001
Level of ID				
Mild	0[a]	-	-	-
Moderate	1.211	1.663	6.779	< .001
Severe	1.844	2.984	13.381	< .001
Profound	1.391	1.595	10.123	.003
Legal representative [b]				
Professional / other	0[a]	-	-	-
Family member	.776	1.215	3.882	.009

aThis group was designated as the reference category.

bThose without a (reported) legal representative were excluded from regression analyses (N=45).

CI, confidence interval.

Reported diagnostics

Genetic testing was reported for 141 individuals: 82 (58.2%) received a test once, 33 (23.4%) twice, 15 (10.6%) three times, 8 (5.7%) four times, and 3 (2.1%) five times , with a total of 248 tests (Table 3). Metabolic testing additional to genetic testing was reported for eight (5.7%) individuals with none having positive metabolic test results, although specification on type of metabolic test was lacking. Mean age at genetic testing was 27.1 (SD 17.8) years old, with information missing for 7 cases. Karyotyping was reported most frequently (n=73) followed by Fragile X syndrome testing (n=49). In total, 51 individuals were reported to have a genetic diagnosis associated with ID, with genetic test results only available for 19 (Supplementary Table B). Twenty-one individuals (13.9%) were reported to have a VUS, and in 15 individuals (10.0%) a clinical diagnosis or genetic variant was mentioned by the care providers as cause of the ID, but insufficiently or incorrectly described in the absence of a letter of a clinical geneticist.

Table 3. Variables on 151 individuals for whom a genetic diagnosis and/or genetic testing results were reported in care files.

Variables	N	%
Genetic testing reported	141	93.4
First genetic test:		
<18 years	57	40.4
≥18 years	74	52.5
Not reported	10	7.1
Last time referred for genetic counseling to geneticist by:		
Intellectual disability physician	38	25.2
Pediatrician	14	9.3
General practitioner	9	6.0
Other medical specialist	4	2.6
Not reported	86	57.0

Of those individuals with genetic test results reported, 57 (40.4%) had their first genetic test during childhood and 74 (52.5%) individuals during adulthood. Individuals who had their first genetic test in childhood received significantly more genetic diagnoses (X^2 (1, N = 141) = 10.137, p = .001). There was no significant difference in ID severity between individuals who had their first genetic test during childhood or adulthood (X^2 (3, N = 140) = 7.434, p = .059).

Reported genetic testing over the years

Over the years, the frequency and types of genetic tests changed (Table 4; Figure 4). Before 2005, mainly karyotyping, fluorescent in-situ hybridization (FISH) and Fragile X testing were performed. From 2005, microarrays were reported. Exome sequencing within this population was reported since 2013, with the exception of one reported in 2008 (possibly incorrect, considering the advent of this technique).

Before 1995, the majority of genetic tests were reported to be performed in children and young adults, while older adults were incidentally tested. In 2004, the first individual over the age of 60 years was tested, and since then 14 more. In the last two decades, older individuals have been increasingly

tested according to care files, and the overall number of genetic tests reported increased over time.

Of 51 individuals who were reported to have received microarray analysis, 7 (13.7 %) received a diagnosis, and another 18 (35.3%) received additional exome sequencing analysis, yielding 3 (6%) additional diagnoses. Of the total number of individuals who received exome sequencing (N=28), 5 (17.9%) had a genetic diagnosis and 11 (39.3%) a VUS.

Table 4. Types of tests reported in medical care files of 141 individuals with ID.

Type of test	Number of tests		Confirmed diagnosis		Age at testing
	N	%	N	%	Mean ± SD
Karyotyping	66	26.6	15	29.4	19.1 ± 17.2
Microarray	51	20.6	7	13.7	28.5 ± 17.9
Fragile X testing	49	19.8	6	11.8	28.3 ±18.3
FISH	13	5.2	5	9.8	25.0 ± 16.4
Exome sequencing	28	11.3	5	9.8	29.4 ± 14.6
MLPA	3	1.2	1	2.0	43.0 ± 2.6
Metabolic	8	3.2	0	0.0	27.3 ± 13.0
Unknown	30	12.1	12	23.5	29.0 ± 21.1
Total	**248**	**100**	**51**	**100**	

FISH, fluorescent in-situ hybridization; ID, intellectual disability; MLPA, multiplex ligation-dependent probe amplification.

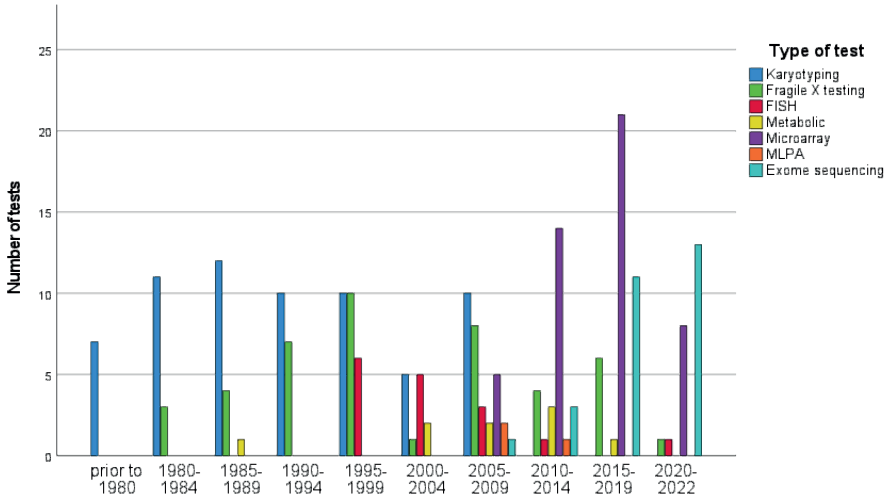


Figure 4. Evolution of different types of tests performed over the years, as reported in files of 141 individuals with ID.

FISH, fluorescent in-situ hybridization; ID, intellectual disability; MLPA, multiplex ligation-dependent probe amplification.

Discussion

This retrospective chart review shows variable reporting of genetic diagnoses by different types of care providers, revealing a gap for optimal personalized care for individuals with ID. Significant associations were found between availability of information on genetic etiology (including genetic testing and diagnoses) in care files and individual's age, level of ID, and the legal representative's relationship to the individual.

Reporting genetic diagnoses in multidisciplinary ID care

Genetic causes of the ID were reported in 90% of medical files, 39% of psychodiagnostic files, and 75% of files used by professional caregivers, although VUS or incorrectly described genetic information were mentioned as cause of the ID in some of these different care files, thus misinterpreting these ambiguous or uncertain genetic findings. This suggests that adequate documentation of a genetic diagnosis is not standard part of

multidisciplinary ID care. Barriers for reporting genetic diagnoses to explain the ID in multidisciplinary care have been found to include lack of awareness of potential benefits, lack of communication and harmonization of coding, and difficulty interpreting the results.[17] It may imply that many individuals with ID miss out on disorder-specific medical care, psychological care, and support. This is unfortunate, as a genetic diagnosis can provide detailed information on the prognosis of the disorder, associated somatic and neuropsychiatric manifestations, and targets for prevention, treatment, and management. It is also important for unaffected family members who might be at risk of passing on a genetic condition to their future children. As it may have impact on all life domains, awareness of all types of HCPs involved is necessitated to improve care.[18]

Surprisingly, not all physicians who were involved in the care team had a reported genetic diagnosis in their medical files, while it was mentioned in one of the other care files, although not verified due to absence of a letter of a clinical geneticist. Coordinating physicians should have direct access to the genetic test results, which means they could inform and update the multidisciplinary team to enable personalized and disorder-specific care, and refer to expert centers where available. From a medical perspective, this may include each body system, including epilepsy management,[19] tumor screening,[15] prevention for sensitivity to obesity[20] and movement disorders,[21] for which a dietician, physiotherapist, or occupational therapist should be involved as well. Without knowledge on the etiology, physicians will not identify and refer candidates who may benefit from disorder-specific care, including condition-specific guidelines or targeted treatments, such as indicated by the Treatable ID (Web) App.[7]

Psychologists and behavioral therapists have a major role with regard to timely consultation of other experts, psychoeducation of care teams and families, treating complex behavioral manifestations, and providing information on appropriate behavioral interventions, guidance and mentoring, preventing frustration, crisis and overmedication.[13] Increasing knowledge on syndrome-specific behavioral manifestations is available.[22,23]

Caregivers are often the first to detect possible disorder-specific manifestations. Understanding the etiology of somatic and behavioral manifestations is of great importance for early signaling and to respond adequately. Especially in complex situations, comprehending the cause and support needs contributes to establishment of a shared concept and vision and multidisciplinary management. This may increase empowerment and anticipatory care planning.

For legal representatives who take care decisions for the affected individual, we found that individuals with close proximity of their legal representative such as first- and second-degree family member appeared to be more likely to receive genetic testing compared to those with a legal representative other than direct family, such as a professional. Family may be more engaged in health management, and may also directly benefit from a diagnosis by better understanding and acceptance, information on recurrence risks and prenatal diagnostic options, and prognostic value about whether someone could still live at home or need professional caregivers. Additional benefits for the affected individual or family members include supportive care, special education or tools, access to expertise centers and (peer) support groups, and financial and emotional support.[24–26]

Diagnostic care gap

If current local and international (pediatric) guidelines were followed, one would expect that most individuals with ID had been referred to a clinical geneticist.[27] However, in our study, only in 40% of individuals with ID reports on genetic etiology were available in care files. A genetic diagnosis was identified in 34% of these individuals, although official results were often not available in the electronic care file system. These results on current clinical practice demonstrate that genetic testing is underutilized, comparable to a previous study in Scotland that reported 41% of individuals with ID had genetic testing with a reported genetic cause for ID in 6%.[28]

We found that more severe levels of ID, lower age, and close proximity of the legal representative's relationship to the individual were associated with increased reporting of information on genetic etiology, indicating disparities in access to genetic testing. Notably, genetic testing in individuals with ID

might differ throughout countries and cultures. Since European countries such as the Netherlands have a high standard with regard to easy and paid access to medical care, the care gap may be expected to be even greater in other countries.

Factors associated with availability of information on genetic etiology in files may indicate both a reporting and diagnostic care gap. Individuals with a higher age appeared to be less likely to have reported information on receiving genetic testing and diagnosis. Our results confirm previous findings that a genetic diagnosis is lacking in many adults,[29,30] possibly including reasons such as less relevance to parents of affected adults in terms of recurrence risk. As the largest population comprises adults, of which the majority did not receive a genetic diagnosis, these might thus miss out on personalized care.

Furthermore, individuals with mild ID appeared to be less likely to have reported information on genetic testing and diagnosis compared to those with moderate, severe or profound ID, possibly due to HCPs being less likely to consider genetic testing for mild ID. Those with mild ID might thus more frequently miss out on disorder-specific care and interventions, underlining the need of awareness and guidelines.

Several barriers for the integration of genetic diagnoses into ID care may exist, including lack of parents for trio exome sequencing, financial issues, and lack of motivation by HCPs.[31] Practical barriers mentioned by physicians in previous research include lack of capacity or unavailability of consent by caregivers, burden and distress, unacknowledging benefits and skepticism about clinical utility especially in adults, and a lack of training resulting in difficulty interpreting and explaining genetic test.[17,28]

Recommendations and future directions

To overcome barriers and contributors to care gaps to identify individuals with ID at risk for underdiagnosis and undertreatment of genetic disorders, recommendations are provided in Table 5. Care organizations should connect with regional clinical genetic centers to reduce the referral threshold and diagnostic delay, and for reanalysis of a VUSses in genes for

which functional tests are available. This may include episignatures which could provide conclusive findings for around 70 known ID syndromes as these have been considered highly sensitive, and specific DNA methylation biomarkers.[32] Education for affected individuals, families, caregivers, (professional) legal representatives, and all types of HCPs on both the importance of genetic testing and the genetic diagnosis may increase awareness and empowerment, and improve quality of multidisciplinary personalized care.[33–35] Adult care, which has usually been variable and fragmented, has greatly improved, advocating for holistic expert care worldwide.

Pediatric guidelines should be extended to adults, since implications of a diagnosis are important for the adult population as well. Individuals with no or borderline ID could also have a genetic disorder, as many genetic syndromes show a great heterogeneity within the disorder, which also requires special attention in psychiatric care. An update on current diagnostic guidelines including genetic testing and counseling in psychiatry has been proposed.[36] Awareness on factors that contribute to a diagnostic care gap should be increased to prevent them from missing out on personalized treatments, management and screening.

Furthermore, electronic care file systems should be improved for this patient population. Protocols should be established for harmonized coding of genetic diagnoses such as using OMIM and ORPHA code. Communication between HCPs should be facilitated by ICT systems, ensuring continuity, transferability and linkage to central relevant sites.

Expertise centers for rare diseases should (inter)nationally assemble, like the European Reference Network (ERN-)ITHACA (https://ern-ithaca.eu/), as these disorders collectively affect many individuals worldwide. These networks contribute to disorder-specific knowledge, including natural course, updating information with regard to the disorder and treatment options, setting up registries, and guidelines, and implementing these in national and regional care networks. This should be performed together with affected individuals and representatives, to also ensure availability of other resources for specific disorders, such as (peer) support groups.[28]

Future research is necessary to examine why knowledge of genetic testing has not been fully implemented, to further identify barriers to personalized care. For instance, as the natural course of (ultra)rare disorders is often unknown, health care providers may question whether a genetic diagnosis really results in better care at present.[37] However, positive experiences in care and benefits for individuals should inspire all to enable and improve disorder-specific care.

Table 5. Barriers and contributors to care gaps to identify individuals with ID at risk of underdiagnosis and undertreatment of genetic disorders. Recommendations are provided by the authors to enable disorder-specific personalized care and empowerment with regard to diagnostics.

	Barriers	Recommendations
	Limited access to genetic testing	Develop, update, and implement (international) protocols and guidelines for genetic testing (especially for adult ID) Stimulate close collaborations between (academic) clinical genetic centers and physicians involved in ID care Facilitate periodic consultations (live or virtual) with a clinical geneticist at the ID facility (e.g., for pre- and post-test counseling, and treatment options) Reduce practical barriers to testing (e.g., train HCPs for genetic diagnostics in regional care networks) Reduce (patient) burden of testing (e.g., using saliva samples (when suitable for the intended test) instead of blood samples) Implement protocol for periodic reanalysis of variants of uncertain significance and repeat genetic testing when no diagnosis was identified Increase transparency on insurance reimbursement of genetic testing if applicable Develop accessible and comprehensible information on somatic and neuropsychiatric manifestations of genomic variants for all HCPs involved Increase understanding of the importance of recurrence risks and prenatal diagnostics for affected individual or (healthy) family members; refer to clinical geneticist in case of unknown diagnoses Increase awareness of the implications of possible negative attitudes towards genetic testing among affected individuals, carers and HCPs (e.g., perceived low yield, insurance problems, fear of stigmatization)
	Limited reporting; coding and harmonization	Establish protocols to harmonize coding and facilitation of ICT systems for communication between HCPs, also to ensure continuity of care

Table 5. continued

	Contributors to decreased reporting of genetic diagnoses	Recommendations
	Type of HCP	Provide education and information to understand importance of a genetic diagnosis for care, for physicians, psychologists, and caregivers
		Improve availability of, and access to, physicians with knowledge on genetic disorders and associated manifestations
		Clarify the role of coordinating physician for referring for (re-)evaluation of genetic diagnosis and inform other care providers
		Implement genetic etiology as standardized part of reporting in medical files and individual support plans in individuals with ID
	Age	Explicitly include adults with ID in guidelines for genetic testing
		Ensure inclusion of genetic test results when transferring individuals with ID to other HCPs, for example in the transition from pediatric to adult care or from parental home to residential care
		Ensure access to all medical information by the coordinating local physician, especially when transitioning to adult care
	Level of ID	Improve awareness of the benefits of genetic testing in care providers of individuals with mild ID and/or limited somatic comorbidity, including guidelines for indications for genetic testing in individuals with for borderline intelligence, e.g., with suspect somatic, psychiatric or neurologic comorbidity
	Legal representative	Increase awareness on care gap of absence of a family member as legal representative, and education for caregivers

HCP, healthcare provider; ID, intellectual disability.

Strengths and limitations

This is the first study that elaborately investigated the integration of genetic diagnoses into multidisciplinary ID care in a large sample based on a sample size calculation. However, representativeness of the data may be affected by the consent procedure: bias might have occurred, since for individuals living within sheltered care facilities of the organization a standard questionnaire is included for providing consent. Selection bias with regard to symptoms, dysmorphisms, suspicion of syndromes, and comorbid features towards those who received genetic testing was not investigated. Moreover, negative genetic test results might not have been documented or (non-digital) information might be lost by switching care

facilities or care providers, such as in the transition from pediatric to adult care. Due to privacy regulations, genetic results are usually only sent to the referring physician. Also, letters from clinical geneticists were often lacking in medical files, and genetic findings were sometimes unclear or incorrectly described by the care provider. Additional genetic variants of clinical relevance were not reported, although these have also been identified in genetic syndromes and ID.[38] As this was a retrospective study in a clinical setting, we could not examine the diagnostic yield of genetic testing.[39]

We encountered difficulties related to inconsistent use of terminology and the lack of a uniform registration in the electronic care system where diagnoses could be found. Genetic disorders are often known by multiple names, possibly resulting in confusion and illustrating the importance of education amongst care providers.

Conclusion

This study showed variable reporting of genetic diagnoses in multidisciplinary ID care files. Type of reporting care provider, milder levels of ID, a higher age, and no family member as legal representative were associated with less reporting and may consequently limit personalized multidisciplinary care. Due to fast advances in the field of diagnostics and targeted interventions, closer collaboration between academia and care organizations is necessary to improve integration of knowledge into daily multidisciplinary practice. Increased genetic testing and adequate reporting of test results over life may improve patient support, outcomes, and allow targeted therapies and surveillance.

Acknowledgements

The authors would like to thank ‘s Heeren Loo and Johan Brands for contributing to the data acquisition.

Funding

The study is financially enabled by the Amsterdam UMC and care organization ‘s Heeren Loo.

Conflict of interest

The authors declare that they have nothing to disclose.

Authors' contributions

AM was involved in the study concept and design, had a major role in the acquisition of data, statistical analyses, interpreting data, and drafting the manuscript. EB was involved in the study concept and design, statistical analyses, interpreting data, and revising the manuscript. SB had a major role in the acquisition and interpretation of data, and drafting the manuscript. CS, LH, MC, BB, and FW substantially contributed to the conception and design of the study, and revised the manuscript. MvH, MA, and CvK were involved in the interpretation of data and revised the manuscript. AvE was involved in the study concept and supervising the design, interpreting data, and drafting and revising the manuscript.

References

1. Pinchefsky E, Shevell M. Intellectual Disabilities and Global Developmental Delay. In: Handbook of DSM-5 Disorders in Children and Adolescents. 2017.
2. American Psychiatric Association. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders Fifth Edition. Arlington. 2013.
3. Pekeles H, Accogli A, Boudrahem-Addour N, Russell L, Parente F, Srour M. Diagnostic Yield of Intellectual Disability Gene Panels. Pediatr Neurol. 2019;92.
4. Anazi S, Maddirevula S, Faqeih E, Alsedairy H, Alzahrani F, Shamseldin HE, et al. Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. Mol Psychiatry. 2017;22(4).
5. Maia N, Nabais Sá MJ, Melo-Pires M, de Brouwer APM, Jorge P. Intellectual disability genomics: current state, pitfalls and future challenges. Vol. 22, BMC Genomics. 2021.
6. Savatt JM, Myers SM. Genetic Testing in Neurodevelopmental Disorders. Vol. 9, Frontiers in Pediatrics. 2021.
7. Hoytema van Konijnenburg EMM, Wortmann SB, Koelewijn MJ, Tseng LA, Houben R, Stöckler-Ipsiroglu S, et al. Treatable inherited metabolic disorders causing intellectual disability: 2021 review and digital app. Vol. 16, Orphanet Journal of Rare Diseases. 2021.
8. Overwater IE, Rietman AB, van Eeghen AM, de Wit MCY. Everolimus for the treatment of refractory seizures associated with tuberous sclerosis complex (TSC): Current perspectives. Therapeutics and Clinical Risk Management. 2019.
9. Van Eeghen AM, Bruining H, Wolf NI, Bergen AA, Houtkooper RH, Van Haelst MM, et al. Personalized medicine for rare neurogenetic disorders: can we make it happen? Cold Spring Harb Mol Case Stud. 2022;8(2):a006200.
10. Coppus AMW. People with intellectual disability: What do we know about adulthood and life expectancy? Vol. 18, Developmental Disabilities Research Reviews. 2013.
11. Both P, ten Holt L, Mous S, Patist J, Rietman A, Dieleman G, et al. Tuberous sclerosis complex: Concerns and needs of patients and parents from the transitional period to adulthood. Epilepsy Behav. 2018.
12. Van Remmerden MC, Hoogland L, Mous SE, Dierckx B, Coesmans M, Moll HA, et al. Growing up with Fragile X Syndrome: Concerns and Care Needs of Young Adult Patients and Their Parents. J Autism Dev Disord. 2020.
13. Davis AS, Phelps LA. Psychoeducational Implications of Neurodevelopmental Genetic Disorders. Sch Psychol Q. 2008;23(2).
14. Egger J, Egger C, Woodcock K, De Witte L, Van Den Bree M, Van Dongen L, et al. Editorial: Treatment of psychopathological and neurocognitive disorders in genetic syndromes: In need of multidisciplinary phenotyping and treatment design. Vol. 13, Frontiers in Psychiatry. 2022.
15. Northrup H, Aronow ME, Bebin EM, Bissler J, Darling TN, de Vries PJ, et al. Updated International Tuberous Sclerosis Complex Diagnostic Criteria and Surveillance and Management Recommendations. Pediatr Neurol. 2021;123.
16. VZinfo. Volksgezondheid en Zorg: Verstandelijke beperking | Leeftijd en geslacht [Internet]. [cited 2023 May 30]. Available from: https://www.vzinfo.nl/verstandelijke-beperking/leeftijd-en-geslacht.
17. Finucane BM, Ledbetter DH, Vorstman JA. Diagnostic genetic testing for neurodevelopmental psychiatric disorders: closing the gap between recommendation and clinical implementation. Vol. 68, Current Opinion in Genetics and Development. 2021.
18. Oliver C, Adams D, Allen D, Bull L, Heald M, Moss J, et al. Causal Models of Clinically Significant Behaviors in Angelman, Cornelia de Lange, Prader-Willi and Smith-Magenis Syndromes. In: International Review of Research in Developmental Disabilities. 2013.
19. Curatolo P, Bjørnvold M, Dill PE, Ferreira JC, Feucht M, Hertzberg C, et al. The Role of mTOR Inhibitors in the Treatment of Patients with Tuberous Sclerosis Complex: Evidence-based and Expert Opinions. Drugs. 2016;76(5).

20. Kaur Y, de Souza RJ, Gibson WT, Meyre D. A systematic review of genetic syndromes with obesity. Vol. 18, Obesity Reviews. 2017.
21. von Scheibler ENMM, van Eeghen AM, de Koning TJ, Kuijf ML, Zinkstok JR, Müller AR, et al. Parkinsonism in Genetic Neurodevelopmental Disorders: A Systematic Review. Vol. 10, Movement Disorders Clinical Practice. 2023.
22. Moskowitz LJ, Jones EA. Uncovering the evidence for behavioral interventions with individuals with fragile X syndrome: A systematic review. Vol. 38, Research in Developmental Disabilities. 2015.
23. De Vries PJ, Whittemore VH, Leclezio L, Byars AW, Dunn D, Ess KC, et al. Tuberous Sclerosis Associated Neuropsychiatric Disorders (TAND) and the TAND checklist. Pediatr Neurol. 2015.
24. de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. N Engl J Med. 2012;367(20).
25. Romano C. The clinical evaluation of patients with mental retardation/intellectual disability. Monogr Hum Genet. 2010;18.
26. Moeschler JB, Shevell M, Schaefer GB, Bull MJ, Enns GM, Gruen JR, et al. Clinical genetic evaluation of the child with mental retardation or developmental delays. Pediatrics. 2006;117(6).
27. Manickam K, McClain MR, Demmer LA, Biswas S, Kearney HM, Malinowski J, et al. Exome and genome sequencing for pediatric patients with congenital anomalies or intellectual disability: an evidence-based clinical guideline of the American College of Medical Genetics and Genomics (ACMG). Genet Med. 2021.
28. De Villiers J, Porteous M. Genetic testing of adults with intellectual disability. Psychiatrist. 2012;36(11).
29. Posey JE, Rosenfeld JA, James RA, Bainbridge M, Niu Z, Wang X, et al. Molecular diagnostic experience of whole-exome sequencing in adult patients. Genet Med. 2016;18(7).
30. Wolfe K, Strydom A, Morrogh D, Carter J, Cutajar P, Eyeoyibo M, et al. Chromosomal microarray testing in adults with intellectual disability presenting with comorbid psychiatric disorders. In: European Journal of Human Genetics. 2016.
31. Rojnueangnit K, Anthanont P, Khetkham T, Puttamanee S, Ittiwut C. Genetic diagnosis for adult patients at a genetic clinic. Cold Spring Harb Mol Case Stud. 2022;8(7):a006235.
32. Sadikovic B, Levy MA, Kerkhof J, Aref-Eshghi E, Schenkel L, Stuart A, et al. Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. Genet Med. 2021;23(6).
33. Peabody J, DeMaria L, Tamandong-LaChica D, Florentino J, Acelajado MC, Burgon T. Low Rates of Genetic Testing in Children With Developmental Delays, Intellectual Disability, and Autism Spectrum Disorders. Glob Pediatr Heal. 2015;2.
34. Wolfe K, Stueber K, McQuillin A, Jichi F, Patch C, Flinter F, et al. Genetic testing in intellectual disability psychiatry: Opinions and practices of UK child and intellectual disability psychiatrists. J Appl Res Intellect Disabil. 2018;31(2).
35. Reiff M, Ross K, Mulchandani S, Propert K, Pyeritz R, Spinner N, et al. Physicians' perspectives on the uncertainties and implications of chromosomal microarray testing of children and families. Clin Genet. 2013;83(1).
36. Pinzón-Espinosa J, van der Horst M, Zinkstok J, Austin J, Aalfs C, Batalla A, et al. Barriers to genetic testing in clinical psychiatry and ways to overcome them: from clinicians' attitudes to sociocultural differences between patients across the globe. Vol. 12, Translational Psychiatry. 2022.
37. Navon D. How do genetic tests answer questions about neurodevelopmental differences? A sociological take. Dev Med Child Neurol. 2022;64(12):1462–9.
38. Girirajan S, Eichler EE. Phenotypic variability and genetic susceptibility to genomic disorders. Hum Mol Genet. 2010;19(R2).
39. Kharazmi AA, Aslani S, Kristiansen MF, Dahl EE, Berner-Hansen M. Indications and diagnostic yield of small-bowel capsule endoscopy in a real-world setting. BMC Gastroenterol. 2020;20(1).

Supplementary materials

Table A. Factors associated with availability of information on genetic etiology in care files using univariate analyses (*N* = 380).

Variable	Effect size	p
Age	r = -0.16	.002
Sex	X^2 = 0.049	0.83
Level of ID	X^2 = 28.898	<.001
Location of receiving care	X^2 = 29.567	.08
Medical care on site	X^2 = 0.349	.55
Legal representative	X^2 = 17.323	<.001

ID, intellectual disability.

Table B. Genetic diagnoses as reported in the electronic care system.

Reported genetic diagnosis	OMIM	N	Genetic test results unavailable
Down syndrome	190685	15	15
Fragile X syndrome	300624	8	7
Smith-Magenis syndrome	182290; 607642	5	3
22q11.2 deletion syndrome	188400; 192430	4	1
Epileptic encephalopathy	176260	2	
Prader-Willi syndrome	176270; 615547	2	1
Pitt-Hopkins syndrome	610954	2	1
18p duplication syndrome	N.A.	1	
18p deletion syndrome	N.A.	1	
ZNF292 syndrome	619188	1	
Miller-Dieker syndrome	247200	1	
1q21.1 microdeletion syndrome	612474	1	
Trisomy 9p	N.A.	1	
PURA syndrome	616158	1	
Alpha-thalassemia-intellectual disability syndrome	141750	1	
Tuberous Sclerosis Complex	191100; 613254	1	1
Schindler disease	609241; 609242	1	1
Williams syndrome	194050	1	1
DeSanto Shinawi syndrome	616708	1	1
WDFY3-related syndrome	N.A.	1	
Total		51	32

To conclude

Chapter 9

General discussion

Targeted therapies are increasingly becoming available for rare genetic neurodevelopmental disorders (RGNDs). The research described in this thesis illustrates the journey from a (genetic) diagnosis toward evidence-based treatments. First, targeted treatments ask for a diagnosis: without a diagnosis, a specific treatment might even not be considered and an affected individual might not be qualified to participate in clinical trials for new treatment options. Early diagnosis may be essential to prevent irreversible damage. Second, the question is whether a treatment is effective to enable evidence-based medicine and guidelines. Third, what do we consider effective? Clinicians and researchers should aim to measure what really matters to affected individuals. And finally, if effective, it could be questioned how the therapy will become accessible to the affected individual.

Personalized care, i.e., the implementation of etiology-driven health monitoring and treatments with the potential of changing the disorder's course,[1] is paramount for improving health and quality of life. However, implementation of therapeutic interventions via clinical trials for patient populations that are rare and heterogeneous faces specific challenges in methodology, outcome measures, and financial, organizational and regulatory barriers. Therefore, a methodological framework and trial service are essential to optimize the potential of personalized care for both the affected individual and the concerning population.

Upcoming interventions and treatment options

The opportunities for personalized and (targeted) therapies are drastically increasing for RGNDs and metabolic disorders. For a growing number of disorders, novel treatment targets are identified in both *in vitro* and *in vivo* studies, paving the way for personalized therapies. Precise genetic diagnoses may result in precise therapeutic approaches. However, translation of novel treatments to care may be cumbersome in terms of access and reimbursement, resulting in an unacceptable situation in which novel therapies do not reach the affected individual with lack of evidence-based care.

Each disorder has its unique underlying cause, mechanism, and symptomatology. This requires targeted treatment approaches, although treatment effects may also be disorder transcending.[2] For instance, a renewed clinical interest in the use of cannabidiol (CBD) has resulted in market approval by the European Medicines Agency (EMA) as add-on treatment for refractory epilepsy in Tuberous Sclerosis Complex (TSC), Dravet syndrome, and Lennox-Gastaut syndrome (*GW Pharma Ltd. Epidyolex [Cannabidiol] Oral Solution*). Promising findings for effectiveness for severe behavioral manifestations in RGNDs and ID have been reported as well.[3–5] As this CBD product will be available due to recent market approval, evidence is needed for the effectiveness and adverse effects when used to treat behavioral manifestations. Therefore, we designed a clinical trial in three different RGNDs, based on availability, urgency and expertise of the centers and patient clinics. The proposed methodology (**Chapter 4**) may also be used for other RGNDs to provide evidence of the effectiveness of CBD. Although the exact mechanism of action of CBD is unclear, several mechanisms are proposed, including a specific mechanism why it might be effective in individuals with Fragile X syndrome.[6,7] If this is indeed the actual working mechanism, it might support accessibility of CBD as a targeted therapy for Fragile X syndrome when positive results are found. However, it is yet unclear whether the medication will be approved and reimbursed for this indication in other individuals with RGNDs or ID of unknown cause not included in this first trial, even though it is not feasible to include all thousands of rare disorders. Without a clearly described mechanism of action, accessibility for those individuals will be hampered.

Research on the efficacy of treatment strategies for RGNDs has been limited. This may result in uncertain efficacy and polypharmacy, which is considered a clinical pitfall in individuals with ID and psychiatric disorders, with the risk of causing iatrogenic comorbidity.[8] There is increasing evidence for differential treatment response and tolerability for regular medication in RGNDs.[9–11] Regular medication might be specifically (in) effective for particular disorders. An example can be found in vigabatrin as a treatment for epilepsy, that is specifically effective in individuals with TSC.[12,13] For attention-deficit/hyperactivity disorder (ADHD), more information is needed about treatment efficacy and side effects in RGNDs.

Therefore, we investigated the effectiveness of methylphenidate in children and adults with Smith-Magenis syndrome (SMS), well-established as first-line pharmacological treatment for ADHD (**Chapter 3**).

Choosing the right trial design

Although new treatment targets are currently identified, lack of evidence may lead to affected individuals missing out on possibly effective interventions. Heterogeneity and small patient populations pose a great challenge to evidence-based medicine.[14] Trial methodology is therefore paramount to generate evidence for RGNDs.

Choice of a single-case experimental designs

Single-case experimental designs (SCEDs), including the N-of-1 design (Figure 1), provide a powerful alternative to the conventional parallel-group randomized controlled trials (RCTs). These RCTs generally assess the average treatment effect, often exclude affected individuals due to strict eligibility criteria such as comorbidities, and cause difficulties with recruitment due to small sample sizes.[1,15–17] SCEDs, however, examine the effect of an intervention for an individual who acts as their own control. Series can be conducted to pool data to draw conclusions about effectiveness for a larger group of individuals. They take into account heterogeneity both within and between participants and may be as robust as large parallel-group RCTs,[18,19] provided that they are properly executed. Under certain conditions, SCEDs can reliably inform personalized evidence-based clinical treatment decisions. These designs are especially suitable for small populations who may also show great inter- and intra-individual variability of the disorder over time, preventing them from missing out on important therapies.

Of the SCEDs, the N-of-1 design is preferred when possible, and might be considered the new 'gold standard' for demonstrating treatment efficacy in affected individuals in clinical practice. They incorporate much of the rigor of large parallel-group RCTs, but are designed for individuals, and a powerful alternative to the group clinical trials by aggregating results from an N-of-1 series.[20] The N-of-1 design has even been elevated to the highest level of evidence for treatment efficacy of an individual with increased

power due to the principle of multiple crossing-overs.[21,22] They address the question of variability in treatment response. If we want to move towards patient-centered care for rare disorders, we should individualize treatment interventions and outcomes in research designs.

N-of-1 studies are especially appropriate in chronic conditions with stable symptoms and when the treatment has a relatively rapid onset (to minimize burden regarding trial duration) and allows an offset with a short half-life (to limit carryover effects).[23] However, the N-of-1 design may be less suitable for certain indications and treatments. For example, for antidepressants and antipsychotics it may take a considerable amount of time before effects become evident. Also, withdrawal effects of the medication can disrupt the findings. Hence, other SCEDs may be recommended in specific situations.[24] The choice for a design may depend, for example, on:

- Type of intervention (e.g., behavioral, symptomatic, disease-modifying);
- The natural course of the construct the intervention is expected to target;
- Time required to enable measuring an effect;
- The 'reversibility' of the effect when an intervention is withdrawn;
- The number of available individuals to participate.

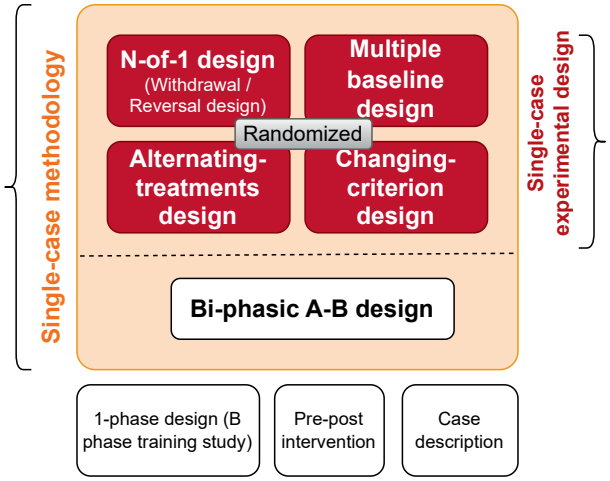


Figure 1. Trial designs especially applicable to rare and heterogeneous patient populations, including single-case experimental designs (SCEDs). Figure adapted from Tate et al. (2013).

Single-case experimental designs in clinical practice

Despite the urgent call for personalized medicine and the challenges the field of rare disorders is facing with regard to novel treatment options and repurposed medicine, SCEDs have not yet been extensively used in either research or clinical practice.[25,26] Researchers might encounter barriers when conducting SCEDs.[27] Recently, a formal International Collaborative Network for N-of-1 Trials and Single-Case Designs (ICN) was established to facilitate wider adoption of SCEDs.[25] In **Chapter 2**, the limited use of N-of-1 studies in RGNDs is described with recommendations to enhance methodological quality and generalizability, feasibility, and personalization, which is underlined by an editorial published in the specific journal in which it was published (*Neurology*).[28] For example, sample size calculations for N-of-1 trials are recommended for generalizability purposes, and statistical analyses could properly address inter- and intra-individual variability.[29] The use of patient registries, (deep) phenotyping, and longitudinal monitoring could help clarify treatment response, natural course and identify possible biomarkers. Future use of the methodological framework should assist clinicians and researchers to realize the sorely needed personalized evidence-based therapies in both care and research settings.

Measuring what matters

The importance of outcome measures

Equally important as the study designs are the outcome measures of choice. By selecting outcome measures, both the patient perspective (including relevancy and burden) and methodological perspectives should be thoroughly considered. The question has frequently been raised whether an intervention is truly inferior to a comparator or whether the selected outcome measurement instrument was inappropriately chosen.[30] This last issue could be related to relevance, feasibility, acceptability, sensitivity in that specific patient population, and other psychometric properties, such as reliability, validity and responsiveness to change. For example in Fragile X syndrome research, substantial efforts were made to evaluate efficacy of promising targeted treatments, but (human) trials did not demonstrate significant benefits based on the primary outcome measure.[31] Despite some promising results on secondary endpoints, these trials were

considered unsuccessful. Negative trial results may lead to companies not being able to continue their drug development program for rare disorders, and obstacles with market authorization or reimbursement of (orphan) drugs.[15,32] Given the high expectations, the negative outcomes have been disheartening to affected individuals and families as they may miss out on important therapies.

Too many, too messy: how to choose outcome measurement instruments?

Deciding upon an appropriate outcome measure can be a daunting task, especially for rare and heterogeneous disorders. Many outcome measures have been applied in previous research and there are plenty of instruments available with each having its own properties. For many RGNDs multiple organ systems are involved and because most existing instruments focus on a single area, researchers have been forced to select a wide variety of tools to measure treatment effects. Additionally, lack of consensus on the most appropriate and relevant outcome measures has resulted in a situation in which several different outcomes have been reported using hundreds of different outcome measurement instruments, which is addressed in **Chapter 5**. For example, cognitive function was measured 333 times with 141 different outcome measurement instruments in clinical trials in RGNDs and ID from the past ten years. Similarly, health-related quality of life was measured 77 times with 24 different instruments, apart from being a broad, abstract, and multidimensional concept that can cover different concepts, obscuring the construct to be measured. Considerations and recommendations have been provided with regard to selecting outcomes and outcome measurement instruments in clinical trials for individuals with RGNDs (**Chapter 5**; Figure 2). The lack of consensus has also resulted in researchers and clinicians designing instruments ad-hoc to capture outcomes relevant to their patients. This is understandable as these might contain less redundant items to be completed and increase acceptability among affected individuals, but might be problematic as well. Measurement properties are often not investigated, and interpretation and extrapolation of results will be hampered.

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

Figure 2. Recommendations with regard to selecting outcomes and outcome measurement instruments in clinical trials for individuals with genetic neurodevelopmental disorders and/or intellectual disability.

ERICA, European Rare Disease Research Coordination and Support Action consortium; NIH-TCB, National Institutes of Health Toolbox Cognitive Battery (NIH-TCB); PROM, patient-reported outcome measure; PROMIS, Patient-Reported Outcomes Measurement Information System.

Challenges with outcome measures in RGNDs and ID

Choosing appropriate outcome measures for RGNDs can be challenging. The heterogeneity and broad range of ID severity in RGNDs may complicate the choice for outcome measures. Hurdles that have been encountered include varying cognitive abilities, high rate of behavioral and emotional disturbances, a lack of stability, practice effects, floor effects, limiting sensitivity, and a lack of consensus on the best measures within a particular construct.[31,33] That has resulted in initiatives, such as the development and validation of the National Institutes of Health Toolbox Cognitive Battery (NIH-TCB) for ID. Furthermore, most existing questionnaires for adults are

often solely available as self-report, while individuals with ID may not be able to self-report. However, outcomes may still be difficult to measure by proxy-reports[34] and other proxy-raters often seem to assess quality of life worse compared to self-raters.[35–40] That urges to find ways to receive information from the individual whom it concerns.[41] It has been suggested that individuals with ID can reliably report on their health status despite of the heterogeneity, provided that instruments are appropriate to their age and cognitive functioning.[42,43] It should be attempted to include the patient's experiences by alternative methods, such as the use of emoticons, experience sampling methods[44] and digital apps.

In search for relevant outcome measurement instruments: disorder-specific and personalized outcome measures

As outcome measurement instruments can be relatively long and time-consuming to fill out, and confronting due to inappropriate questions especially for lower levels of functioning, acceptability is often considered poor.[45] Also, the burden it may pose to the often already overworked caregivers should be taken into account. Furthermore, an effect may be statistically significant, but not clinically or socially relevant, or vice versa. Generally, commonly used questionnaires do not include specific domains of interest and disorder-specific symptoms that are important to the patient populations, due to the heterogeneity and complexity of manifestations. For these target populations, these instruments may thus have limited relevancy and sensitivity.[45–49] Therefore, the desire for **disorder-specific outcome measures** has been expressed by several patient communities to measure the impact of the particular disease on health. In **Chapter 6**, a TSC-specific patient-reported outcome measure (TSC-PROM) is developed and validated together with patient representatives and TSC experts. The TSC-PROM is developed to measure the impact of TSC on mental functioning, physical functioning, activities and participation, social support someone receives and quality of life, using the International Classification of Functioning and Disability (ICF) (World Health Organization 2001). The ICF provides a framework to select a combination of outcome measures that capture all components to better understand the impact of a diagnosis on all life domains, providing relevant interventions for optimal quality of life. PROMs provide quantitative evaluation of symptoms and functioning from

the patient perspective, and have become important for shared-decision making and value-based healthcare.[50] PROMs can be used for monitoring and informing care. It may improve communication between patient and clinician and about treatment outcomes, resulting in increased patient satisfaction with care.

Noticeable, it is not feasible and desirable to develop and validate disorder-specific outcome measures for all thousands of rare disorders, as there is already an overgrowth of outcome measurement instruments, complicating generalizability. However, the tendency towards development of disorder-specific outcome measures and questionnaires designed ad hoc for a trial demonstrates the struggle researchers may experience: lack of an appropriate way to measure what is relevant to affected individuals.

One way to measure what matters to affected individuals is the use of **personalized outcome measures** like the personal questionnaire and Goal Attainment Scaling (GAS),[51] which is also used in the clinical trials explained in **Chapter 3 and 4**. These instruments are generally intended for standardized evaluation of an intervention's effectiveness based on individualized goals. Goals could be different for each patient, but it has to do something with the construct that the treatment will target. In this way, outcomes can be measured that are specifically relevant to the affected individual, which eventually enhances treatment adherences as well.[52] Treatment goals often differ which might unintentionally exclude affected individuals from clinical trials as frequently used outcome measures might not be relevant. This type of instrument is especially meaningful for heterogeneous and small populations, although it has not yet been validated for these populations.

PROMs, GAS or experience sampling methods may thus be considered, enabling quantitative expression of meaningful subjective patient experiences while translating these into evidence. In care setting, each additional outcome measure could be considered a burden, whereas in research setting we should gain knowledge and explore, since there is a lot unknown yet . However, we should still pursue to minimize burden as much as possible. This could be discussed with patient (representatives) as well. The patient's perspective should be included as much as possible by

tailoring the method of measurement to the individual's level of functioning. Also, including ID- or proxy-friendly assessment tools will ensure trial compliance.[1] This personalized approach has the potential of maximizing treatment and trial adherence that is both evidence-based and patient-centered.

Ethical considerations

Several ethical considerations exist with regard to interventions for RGNDs. First of all, individuals without a diagnosis (yet) should receive the same expert care that is increasingly available for specific disorders.[53] As described in **Chapter 8**, less than half of individuals with ID received genetic testing as reported in care files, while a genetic etiology can already be identified in up to 50%.[54–59] There are also health disparities in access to genetic testing, and some were identified in **Chapter 8**. These include age, severity of ID, and the legal representative's relationship to the individual. These sub-groups within RGNDs and ID appeared to be less likely to receive genetic testing. Genetic diagnostics should become accessible and considered for everyone with ID, also for lower-income countries. In the Netherlands, genetic testing is considered an integral part of regular care in children with ID. Due to improved (syndrome-specific) care, life expectancy has increased and the largest population with ID is now adult.[60–62] However, many adults with ID have never been genetically tested, possibly missing out on the advantages of disorder-specific care,[63,64] even though it is yet unclear whether a known genetic etiology truly reaches the care team with implications for management and multidisciplinary disorder-specific care.

Another issue includes access to interventions when effective, but not yet registered for that indication or reimbursement. In a single N-of-1 trial with CBD for a male with Fragile X syndrome, we found substantial improvements on several outcome measures, including GAS. Despite the efforts to substantiate the effectiveness of CBD for this individual and the calculation demonstrating reduced costs in the end, the individual's healthcare insurance did not want to reimburse on a single basis. Parents faced a dilemma: is it worth to pay a few hundred euros per month for these improvements while it is actually not financially viable in their current situation? These issues should be thoroughly discussed in advance when

starting a trial with uncertain possibilities for reimbursement to avoid undesirable and unethical situations.

N-of-1 in clinic or research?

A practical flowchart based on an ethical framework was designed to help distinguish an N-of-1 trial to be considered as part of evidence-based clinical care or representing medical research with need for approval from the institutional review board (IRB).[65] If there is a treatment option and the effectiveness will be examined for the affected individual with the highest level of evidence, one wants to start an N-of-1 trial. Such a trial in one individual is not subject to the Dutch *Medical Research Involving Human Subjects Act* (WMO), because a 'case study' (N=1) has not been considered as research with generalizable results applicable to the population other than the direct population of interest. It is thus considered experimental care. It enables rapid care for a particular affected individual, which is essential to this patient population. If there is the idea to follow the study protocol in more than one individual, possibly and eventually attempting to provide treatment effect estimates at a group level, it is subject to the WMO and often concerns drug research.

From two sides this might be odd and sometimes even undesirable. On the one hand, even for one individual, a medical ethical view may be valuable. Interventions for this population may often be experimental, expensive and burdensome. As such, it might be desirable to have the ethical considerations co-assessed by a recognized and knowledgeable ethical committee, such as a medical ethical committee (METC) or the Dutch Central Committee on Research Involving Human Subjects (CCMO). However, submitting a study protocol including all documentation required for clinical trials with medicinal products (EU Clinical Trial Regulation) is time and effort-consuming. Especially if the METC would assess all single N-of-1 trials, the amount of work for both METC and clinicians may be prohibitive. On the other hand, for N-of-1 trials in a few individuals, starting a clinical trial is comparable with large clinical trials in terms of financial and time effort. For a particular affected individual participating in such a (prospective) clinical trial, it may take at least one year before the trial will start, while health conditions are often devastating with an urgent need for

treatment. These (financial) aspects could even impede research efforts aimed at investigating rare disorders, whereas the affected individuals might be in desperate need.

Potential barriers for N-of-1 trials being subjected to WMO (drug research) may thus include:

- Time and effort-consuming (for both clinicians, researchers, and METC);
- Delay before trial commences;
- Financial aspects (e.g., study medication including placebo, pharmacy costs (including staff costs, quality control, labelling, handling, storage, and dispensing), costs medical ethical review, and monitoring);
- Ambiguity about ethical procedures (e.g., prospective research compared to retrospectively pooling data for treatment effect estimates when same procedure is replicated in multiple individuals).

Toward a rare disorder trial framework

Is it possible to work towards creating a situation in which multiple interventions for several RGNDs could be systematically investigated, using a master protocol?[66,67] A standardized template of approved trial designs and outcome measures, with and without the use of placebo, could be developed in the context of good clinical practice and a trial service. Moreover, deviations of trial designs or (more invasive) outcome measures, and experimental therapies could be periodically presented to a delegation of a medical ethical committee to ensure we are behaving in an ethically responsible manner regarding the vulnerable affected individuals and novel therapies.

Challenges

Trial development for rare disorders faces many challenges that should be addressed to efficiently realize interventional research for this patient population. For investigator-initiated studies (i.e., not initiated by a pharmaceutical company), financial issues may obstruct initiating a trial. In

addition, the administrative work related to initiation of a trial may all lead to a long delay before including the first patient. How should we finance that?

The trial design together with smartly chosen outcome measures might help. Despite requiring fewer participants, financial expenses are likely to remain quite similar in practice, owing to mandatory procedures for clinical trials involving medicinal products. The platform 'Medicijn voor de Maatschappij' (Medicine for Society) is a great initiative for sustainable and affordable availability of medicines for rare diseases. They are a knowledge hub and carry out projects to ensure availability of specific medicines. They assist in exploring the possibilities of conducting clinical trials with medicinal products (e.g., by pharmacy compounding or magistral preparation, or collaborations with pharmaceutical companies). Collaborating with pharmaceutical companies could accelerate accessibility and maintain affordability for academic research through the provision of the investigational drug. However, agreements should be carefully established in conjunction with legal research support. To our experience, arranging contracts might be a time-consuming process.

Enabling 'care for rare' is a matter of prioritization. Opting for large trials for common diseases is an important but straightforward approach, but it may leave several patients with rare complex disorders untreated. Concurrently, the core values of an academic medical center include taking responsibility to care for all individuals. For rare complex disorders, it requires innovation and necessitates financial investment. If the academic hospitals place significance on 'care for rare disorders' – and I believe they should – the hospitals should further facilitate and support this kind of research, to ensure a reasonable timeframe for delivering therapy to those in need.

FUTURE PERSPECTIVES

From clinic to trial to care

To enable evidence-based medicine for individuals with RGNDs, an efficient framework is necessary to get therapies and trial results to the affected individual. A **clinical trial service** can be valuable to catalyze interventional research in rare disorders (Figure 3). A very promising initiative is the Emma Center for Personalized Medicine (www.emmacenter.nl). This dedicated personalized medicine center unites expertise and efforts regarding translational research and care activities, aimed at providing accurate diagnoses, counseling, effective interventions, accessibility, personalized care and eventually prevention. **Master protocols**, referring to an overarching trial design developed to evaluate multiple hypotheses, should be developed for fast access to therapies and to reduce costs as much as possible. These include basket trials (i.e., a targeted therapy is evaluated for multiple diseases that share common molecular alterations or risk factors) and umbrella trials (i.e., multiple targeted therapies are evaluated for a single disease that is stratified into multiple subgroups based on different molecular or predictive risk factors).[67] These may improve efficiency in drug development by diminishing redundancy in trial implementation, enhancing recruitment, sharing control groups, and using biomarkers that are relevant to the intervention's mechanism of action across RGNDs.[66] We should utilize **templates** for trial development and contracts rather than constantly reinventing the wheel, not only for research but also for integration in care.

Furthermore, **collaboration** is key. There are many steps needed for treatment success which can only be achieved by uniting expertise. A trial service is well-positioned to collaborate with METCs, hospital pharmacists or pharmaceutical companies, health insurances and regulatory authorities for accessibility. But also, international collaboration is essential for research in rare disorders, because for some disorders there are only a few affected individuals per country. Such an initiative is the 'International Collaborative Network (ICN) for N-of-1 Trials and Single-Case Designs' (www.nof1sced.org) which is uniquely positioned to facilitate a range of activities to promote, support and advance the use of personalized clinical studies. In this way, personalized trials will become an integral part of clinical practice and health

research by sharing relevant knowledge, experience, expertise, resources, and data through the global partnership between researchers, healthcare providers, and affected individuals. Furthermore, projects such as Solve-RD, the International Rare Diseases Research Consortium (IRDiRC), European Reference Networks (ERNs), and the digital knowledgebases including Treatabolome and Treatable ID (www.treatable-id.org) are very promising in aiming to enable all individuals with a rare disorder to receive accurate diagnosis, care, and available therapy, and shorten the path to therapy by providing access to treatment information for specific disorders.[68–70]

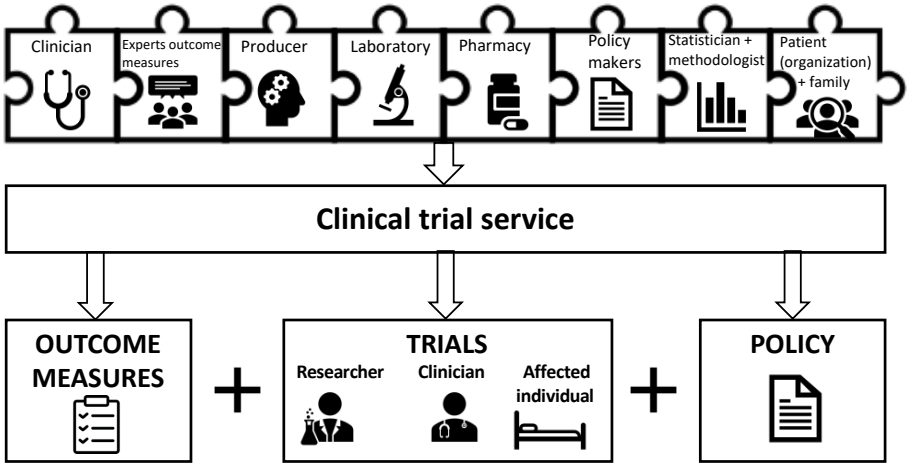


Figure 3. Example of a clinical trial service for rare disorders and required collaborations.

The future landscape of outcome measures

Procedures for measuring outcomes should strive for optimal patient relevance with minimally invasive methods. We should aim to engage the included patients themselves by using applicable methods, because they are the ones who truly experience their feelings with indications of discrepancies between self and proxy-raters. Simultaneously, caregivers, who have often already experienced much burden, might be slightly relieved. Furthermore, we should always question ourselves to what extent each outcome measure is relevant to the affected individual. Including personalized outcomes, such as identified by using PROMs and Goal Attainment Scaling, will contribute to the relevancy and will improve communication and shared decision-making.

Specifically, as there is an overgrowth of available outcome measurement instruments that are not always relevant to all individuals with RGNDs and ID, and development of disorder-specific outcome measures for thousands of disorders is not feasible and desirable, a core outcome set should be developed which can be extended with disorder-specific modules. Also, recent methodological innovations, such as item response theory (IRT), allow PROMs with good measurement properties to be applied across different health conditions, such as the Patient-Reported Outcomes Measurement Information System (PROMIS®).[71–73] IRT can be used to create item banks, which are large sets of calibrated questions measuring a same construct. It enables efficient measurement through short forms or computerized adaptive testing (CAT) to minimize redundant items for specific affected individuals and increase relevancy and efficiency.[74–77] These methods should be more frequently included in interventional research, and the use of ID user-friendly mobile apps in terms of experience sampling methods might be considered as well to track health and improve inclusivity and to know as good as possible the feelings and experiences of the concerned individual, to enable the best personalized care.

Patient involvement

Most importantly, it is vital that the trial and outcome measures capture as much as possible to real-life impact of a disease at an individual level, which is only feasible when the patient's perspective is included. Therefore, affected individuals should be involved in the decision how to measure the effectiveness of an intervention, which is of pivotal importance for trial success. Trials are ultimately aimed at improving patients' well-being. It is thus remarkable that patient involvement was mentioned in only 2% of the clinical trials included in the scoping review about outcome measures (see **Chapter 5**). Moreover, in a clinical trial service, different options for trial designs could be presented to affected individuals, such as a blinded or unblinded A-B design with or without placebo and the N-of-1 design with or without interim analyses (Figure 4). Based on the explained advantages and disadvantages, including the burden and level of evidence, affected individuals, parents and/or caregivers might express their preferences for a design. To foster recruitment and treatment adherence, patient involvement

in the design, intervention, and outcome measures will greatly contribute to the enthusiasm of participants and will increase experienced relevancy.[52]

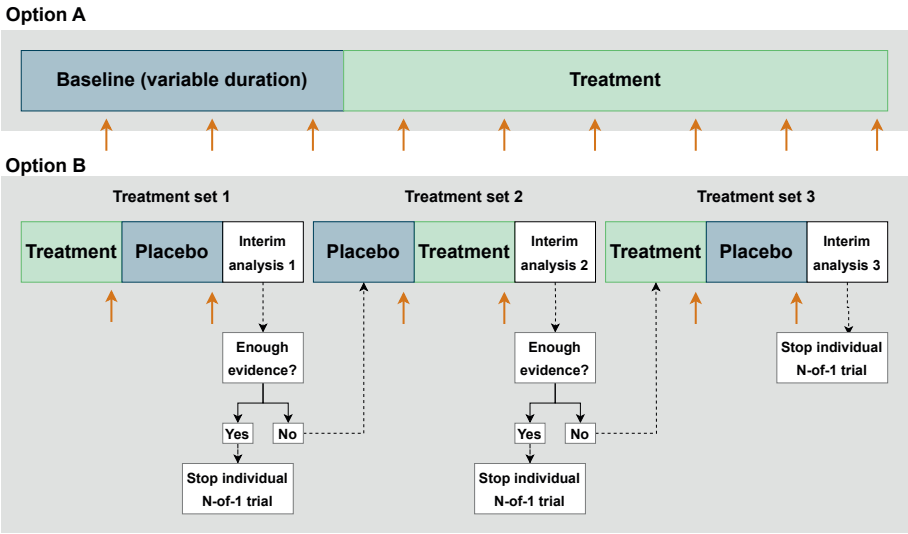


Figure 4. Possible options for trial designs to be presented to affected individuals or family members in a clinical trial service. Preferences should be discussed with regard to level of evidence, (un) blinded periods, the use of placebo, and interim analyses. Orange arrows indicate measurement of outcomes.

Evidence-based care for individuals with RGNDs is challenging, but can be realized. It requires an ethical and methodological framework for trial designs as well as the selection of outcome measures that are relevant and measure what matters. A genetic diagnosis for individuals with ID may provide several benefits, including disorder-specific care, which is important for all involved healthcare professionals. Collaboration is key. Together with affected individuals, caregivers, clinicians, researchers, pharmacists, policy makers, and other experts, we can enable personalized, disorder-specific care for these vulnerable individuals.

References

1. Van Eeghen AM, Bruining H, Wolf NI, Bergen AA, Houtkooper RH, Van Haelst MM, et al. Personalized medicine for rare neurogenetic disorders: can we make it happen? Cold Spring Harb Mol Case Stud. 2022;8(2):a006200.
2. Strydom A, Coppus A, Blesa R, Danek A, Fortea J, Hardy J, et al. Alzheimer's disease in Down syndrome: An overlooked population for prevention trials. Vol. 4, Alzheimer's and Dementia: Translational Research and Clinical Interventions. 2018.
3. Efron D, Taylor K, Payne JM, Freeman JL, Cranswick N, Mulraney M, et al. Does cannabidiol reduce severe behavioural problems in children with intellectual disability? Study protocol for a pilot single-site phase I/II randomised placebo controlled trial. BMJ Open. 2020.
4. Heussler H, Cohen J, Silove N, Tich N, Bonn-Miller MO, Du W, et al. A phase 1/2, open-label assessment of the safety, tolerability, and efficacy of transdermal cannabidiol (ZYN002) for the treatment of pediatric fragile X syndrome. J Neurodev Disord. 2019.
5. Tartaglia N, Bonn-Miller M, Hagerman R. Treatment of Fragile X Syndrome with Cannabidiol: A Case Series Study and Brief Review of the Literature. Cannabis and Cannabinoid Research. 2019.
6. Jung KM, Sepers M, Henstridge CM, Lassalle O, Neuhofer D, Martin H, et al. Uncoupling of the endocannabinoid signalling complex in a mouse model of fragile X syndrome. Nat Commun. 2012;3.
7. Palumbo JM, Thomas BF, Budimirovic D, Siegel S, Tassone F, Hagerman R, et al. Role of the endocannabinoid system in fragile X syndrome: potential mechanisms for benefit from cannabidiol treatment. J Neurodev Disord. 2023;15(1):1–10.
8. De Kuijper G, Evenhuis H, Minderaa RB, Hoekstra PJ. Effects of controlled discontinuation of long-term used antipsychotics for behavioural symptoms in individuals with intellectual disability. J Intellect Disabil Res. 2014.
9. Cheon KA, Ryu YH, Kim JW, Cho DY. The homozygosity for 10-repeat allele at dopamine transporter gene and dopamine transporter density in Korean children with attention deficit hyperactivity disorder: Relating to treatment response to methylphenidate. Eur Neuropsychopharmacol. 2005;15(1).
10. Laje G, Bernert R, Morse R, Pao M, Smith ACM. Pharmacological treatment of disruptive behavior in Smith-Magenis syndrome. Am J Med Genet Part C Semin Med Genet. 2010.
11. van der Poest Clement E, Jansen FE, Braun KPJ, Peters JM. Update on Drug Management of Refractory Epilepsy in Tuberous Sclerosis Complex. Pediatric Drugs. 2020.
12. Curatolo P, Bjørnvold M, Dill PE, Ferreira JC, Feucht M, Hertzberg C, et al. The Role of mTOR Inhibitors in the Treatment of Patients with Tuberous Sclerosis Complex: Evidence-based and Expert Opinions. Drugs. 2016;76(5).
13. Morrissey H, Pradhan T. Escitalopram in the treatment of anxiety in chromosome 18p deletion. Vol. 28, Journal of Child and Adolescent Psychopharmacology. 2018.
14. Stockler-Ipsiroglu S, Potter BK, Yuskiv N, Tingley K, Patterson M, van Karnebeek C. Developments in evidence creation for treatments of inborn errors of metabolism. Vol. 44, Journal of Inherited Metabolic Disease. 2021.
15. Erickson CA, Davenport MH, Schaefer TL, Wink LK, Pedapati E V., Sweeney JA, et al. Fragile X targeted pharmacotherapy: Lessons learned and future directions. Vol. 9, Journal of Neurodevelopmental Disorders. 2017.
16. Overwater IE, Rietman AB, Mous SE, Bindels-De Heus K, Rizopoulos D, Ten Hoopen LW, et al. A randomized controlled trial with everolimus for IQ and autism in tuberous sclerosis complex. Neurology. 2019.
17. Tate RL, Perdices M. Single-Case Experimental Designs for Clinical Research and Neurorehabilitation Settings. Single-Case Experimental Designs for Clinical Research and Neurorehabilitation Settings. 2019.

18. 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.
19. Mitchell GK, Hardy JR, Nikles CJ, Carmont SAS, Senior HE, Schluter PJ, et al. The effect of methylphenidate on fatigue in advanced cancer: An aggregated N-of-1 trial. J Pain Symptom Manage. 2015.
20. Stunnenberg BC, Raaphorst J, Groenewoud HM, Statland JM, Griggs RC, Woertman W, et al. Effect of Mexiletine on Muscle Stiffness in Patients with Nondystrophic Myotonia Evaluated Using Aggregated N-of-1 Trials. JAMA - J Am Med Assoc. 2018;320(22).
21. OCEBM Levels of Evidence Working Group, Durieux N, Pasleau F, Howick J. The Oxford 2011 Levels of Evidence. Group. 2011.
22. OCEBM Levels of Evidence Working Group, Jeremy Howick, Iain Chalmers (James Lind Library), Paul Glasziou, Trish Greenhalgh, Carl Heneghan, Alessandro Liberati, Ivan Moschetti, Bob Phillips, Hazel Thornton OG and MH. The Oxford levels of Evidence 2. Oxford Cent Evidence-Based Med. 2011.
23. Stunnenberg BC, Berends J, Griggs RC, Statland J, Drost G, Nikles J, et al. N-of-1 Trials in Neurology A Systematic Review. Neurology. 2022;98(2).
24. Kazdin AE. Single-case experimental designs. Evaluating interventions in research and clinical practice. Behav Res Ther. 2019;117.
25. Nikles J, Onghena P, Vlaeyen JWS, Wicksell RK, Simons LE, McGree JM, et al. Establishment of an International Collaborative Network for N-of-1 Trials and Single-Case Designs. Contemp Clin Trials Commun. 2021;23.
26. Müller AR, Brands MM, van de Ven PM, Roes KC, Cornel MC, van Karnebeek CD, et al. The Power of 1: Systematic Review of N-of-1 Studies in Rare Genetic Neurodevelopmental Disorders. Neurology. 2021.
27. Kravitz RL, Duan N, Niedzinski EJ, Hay MC, Subramanian SK, Weisner TS. What ever happened to N-of-1 trials? Insiders' perspectives and a look to the future. Milbank Q. 2008.
28. Hartman AL. N-of-1 Trials in Rare Genetic Neurodevelopmental Disorders: Opportunities for Improvement. Vol. 96, Neurology. 2021.
29. Senn S. Mastering variation: Variance components and personalised medicine. Stat Med. 2016.
30. Budimirovic DB, Berry-Kravis E, Erickson CA, Hall SS, Hessl D, Reiss AL, et al. Updated report on tools to measure outcomes of clinical trials in fragile X syndrome. Journal of Neurodevelopmental Disorders. 2017.
31. Hessl D, Sansone SM, Berry-Kravis E, Riley K, Widaman KF, Abbeduto L, et al. The NIH Toolbox Cognitive Battery for intellectual disabilities: Three preliminary studies and future directions. J Neurodev Disord. 2016;8(1).
32. Overwater IE, Rietman AB, van Eeghen AM, de Wit MCY. Everolimus for the treatment of refractory seizures associated with tuberous sclerosis complex (TSC): Current perspectives. Therapeutics and Clinical Risk Management. 2019.
33. Shields RH, Kaat AJ, McKenzie FJ, Drayton A, Sansone SM, Coleman J, et al. Validation of the NIH Toolbox Cognitive Battery in intellectual disability. Neurology. 2020.
34. Ediebah DE, Reijneveld JC, Taphoorn MJB, Coens C, Zikos E, Aaronson NK, et al. Impact of neurocognitive deficits on patient–proxy agreement regarding health-related quality of life in low-grade glioma patients. Qual Life Res. 2017;26(4).
35. Alcantara J, Ohm J, Alcantara J. Comparison of pediatric self reports and parent proxy reports utilizing PROMIS: Results from a chiropractic practice-based research network. Complement Ther Clin Pract. 2017;29.
36. Burks HB, Des Bordes JKA, Chadha R, Holmes HM, Rianon NJ. Quality of life assessment in older adults with dementia: A systematic review. Vol. 50, Dementia and Geriatric Cognitive Disorders. 2021.
37. Griffiths AW, Smith SJ, Martin A, Meads D, Kelley R, Surr CA. Exploring self-report and proxy-report quality-of-life measures for people living with dementia in care homes. Qual Life Res. 2020;29(2).

38. Janse AJ, Gemke RJBJ, Uiterwaal CSPM, Van Der Tweel I, Kimpen JLL, Sinnema G. Quality of life: Patients and doctors don't always agree: A meta-analysis. J Clin Epidemiol. 2004;57(7).
39. Lunsky Y, Bramston P. A preliminary study of perceived stress in adults with intellectual disabilities according to self-report and informant ratings. J Intellect Dev Disabil. 2006;31(1).
40. Morrow AM, Hayen A, Quine S, Scheinberg A, Craig JC. A comparison of doctors', parents' and children's reports of health states and health-related quality of life in children with chronic conditions. Child Care Health Dev. 2012;38(2).
41. Maguire S, Davison J, McLaughlin M, Simms V, Bunting B. Exploring the psychometric properties of self-report instruments used to measure health-related quality of life and subjective wellbeing of adolescents with intellectual disabilities: A Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) systematic review. J Appl Res Intellect Disabil [Internet]. 2023 Apr 26; Available from: http://www.ncbi.nlm.nih.gov/pubmed/37101341.
42. Davison J, Maguire S, McLaughlin M, Simms V. Involving adolescents with intellectual disability in the adaptation of self-reported subjective well-being measures: participatory research and methodological considerations. J Intellect Disabil Res. 2022;66(7).
43. Ingerski LM, Modi AC, Hood KK, Pai AL, Zeller M, Piazza-Waggoner C, et al. Health-Related Quality of Life Across Pediatric Chronic Conditions. J Pediatr. 2010;156(4).
44. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. Depress Anxiety. 2017;34(6).
45. Kang S, Jones A, Shaffer RC, Erickson CA, Schmitt LM. Developing Improved Outcome Measures in FXS: Key Stakeholder Feedback. SSRN Electron J. 2023.
46. Bijkerk CJ, De Wit NJ, Muris JWM, Jones RH, Knottnerus JA, Hoes AW. Outcome measures in irritable bowel syndrome: Comparison of psychometric and methodological characteristics. Am J Gastroenterol. 2003;98(1).
47. Both P, ten Holt L, Mous S, Patist J, Rietman A, Dieleman G, et al. Tuberous sclerosis complex: Concerns and needs of patients and parents from the transitional period to adulthood. Epilepsy Behav. 2018.
48. Krueger DA, Wilfong AA, Mays M, Talley CM, Agricola K, Tudor C, et al. Long-term treatment of epilepsy with everolimus in tuberous sclerosis. Neurology. 2016;87(23).
49. Lee EH, Kwon O, Hahm KB, Kim WH, Kim J Il, Cheung DY, et al. Irritable bowel syndrome-specific health-related quality of life instrument: Development and psychometric evaluation. Health Qual Life Outcomes. 2016;14(1).
50. Gibbons C, Porter I, Gonçalves-Bradley DC, Stoilov S, Ricci-Cabello I, Tsangaris E, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. Vol. 2021, Cochrane Database of Systematic Reviews. 2021.
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.
52. Gaasterland CMW, Jansen-van der Weide MC, Vroom E, Leeson-Beevers K, Kaatee M, Kaczmarek R, et al. The POWER-tool: Recommendations for involving patient representatives in choosing relevant outcome measures during rare disease clinical trial design. Health Policy (New York). 2018.
53. Mueller T, Jerrentrup A, Bauer MJ, Fritsch HW, Schaefer JR. Characteristics of patients contacting a center for undiagnosed and rare diseases. Orphanet J Rare Dis. 2016;11(1).
54. Anazi S, Maddirevula S, Faqeih E, Alsedairy H, Alzahrani F, Shamseldin HE, et al. Clinical genomics expands the morbid genome of intellectual disability and offers a high diagnostic yield. Mol Psychiatry. 2017;22(4).
55. Chiurazzi P, Pirozzi F. Advances in understanding - genetic basis of intellectual disability. F1000Research. 2016;5.
56. Maia N, Nabais Sá MJ, Melo-Pires M, de Brouwer APM, Jorge P. Intellectual disability genomics: current state, pitfalls and future challenges. Vol. 22, BMC Genomics. 2021.

57. Pekeles H, Accogli A, Boudrahem-Addour N, Russell L, Parente F, Srour M. Diagnostic Yield of Intellectual Disability Gene Panels. Pediatr Neurol. 2019;92.
58. Stefanski A, Calle-López Y, Leu C, Pérez-Palma E, Pestana-Knight E, Lal D. Clinical sequencing yield in epilepsy, autism spectrum disorder, and intellectual disability: A systematic review and meta-analysis. Epilepsia. 2021;62(1).
59. Wang J, Wang Y, Wang L, Chen WY, Sheng M. The diagnostic yield of intellectual disability: Combined whole genome low-coverage sequencing and medical exome sequencing. BMC Med Genomics. 2020;13(1).
60. Coppus AMW. People with intellectual disability: What do we know about adulthood and life expectancy? Vol. 18, Developmental Disabilities Research Reviews. 2013.
61. Stepien KM, Kieć-Wilk B, Lampe C, Tangeraas T, Cefalo G, Belmatoug N, et al. Challenges in Transition From Childhood to Adulthood Care in Rare Metabolic Diseases: Results From the First Multi-Center European Survey. Front Med. 2021;8.
62. Presson AP, Partyka G, Jensen KM, Devine OJ, Rasmussen SA, McCabe LL, et al. Current estimate of down syndrome population prevalence in the United States. J Pediatr. 2013;163(4).
63. Posey JE, Rosenfeld JA, James RA, Bainbridge M, Niu Z, Wang X, et al. Molecular diagnostic experience of whole-exome sequencing in adult patients. Genet Med. 2016;18(7).
64. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, et al. Molecular findings among patients referred for clinical whole-exome sequencing. JAMA - J Am Med Assoc. 2014;312(18).
65. Stunnenberg BC, Deinum J, Nijenhuis T, Huysmans F, van der Wilt GJ, van Engelen BGM, et al. N-of-1 Trials: Evidence-Based Clinical Care or Medical Research that Requires IRB Approval? A Practical Flowchart Based on an Ethical Framework. Healthcare. 2020.
66. Cummings J, Montes A, Kamboj S, Cacho JF. The role of basket trials in drug development for neurodegenerative disorders. Vol. 14, Alzheimer's Research and Therapy. 2022.
67. Park JJH, Hsu G, Siden EG, Thorlund K, Mills EJ. An overview of precision oncology basket and umbrella trials for clinicians. CA Cancer J Clin. 2020;70(2).
68. Atalaia A, Thompson R, Corvo A, Carmody L, Piscia D, Matalonga L, et al. A guide to writing systematic reviews of rare disease treatments to generate FAIR-compliant datasets: Building a Treatabolome. Orphanet J Rare Dis. 2020;15(1).
69. Bonne G. The Treatabolome, an emerging concept. Vol. 8, Journal of Neuromuscular Diseases. 2021.
70. Van Karnebeek CDM, Houben RFA, Lafek M, Giannasi W, Stockler S. The treatable intellectual disability APP www.treatable-id.org: A digital tool to enhance diagnosis & care for rare diseases. Orphanet J Rare Dis. 2012;7(1).
71. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol. 2010;63(11).
72. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH roadmap cooperative group during its first two years. Med Care. 2007;45(5 SUPPL. 1).
73. Nguyen TH, Han HR, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. Vol. 7, Patient. 2014.
74. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: Item banking, tailored short-forms, and computerized adaptive assessment. In: Quality of Life Research. 2007.
75. Chang CH. Patient-reported outcomes measurement and management with innovative methodologies and technologies. In: Quality of Life Research. 2007.
76. Khanna D, Krishnan E, Dewitt EM, Khanna PP, Spiegel B, Hays RD. The future of measuring patient-reported outcomes in rheumatology: Patient-Reported Outcomes Measurement Information System (PROMIS). Arthritis Care Res. 2011;63(SUPPL. 11).
77. Thomas ML. Advances in applications of item response theory to clinical assessment. Psychol Assess. 2019;31(12).

Chapter 10

Summary

Millions of people worldwide are affected by a rare genetic neurodevelopmental disorder (RGND). Although individually rare, together these are collectively common, affecting 1-3% of the population. RGNDs are often characterized by complex, multiorgan comorbidity, with neuropsychiatric manifestations and intellectual disability (ID) posing the greatest burden. The lifelong need for care presents challenges for health care professionals when it comes to delivering optimal personalized care, which is the implementation of etiology-driven health monitoring and treatments. This necessitates further research on evidence-based care to improve quality of life of these vulnerable individuals.

Advances in translational research has resulted in emerging development of (targeted) therapies, varying from diets, drug-repurposing, to enzyme and gene therapy. Throughout life, many interventions may be applied in individuals with RGNDs and ID, with risk for over- and/or undertreatment. However, evidence for efficacy is often low due to methodological and statistical challenges in these populations, resulting in affected individuals missing out on possibly effective (targeted) treatments.

Interventional research is challenging due to the rarity and heterogeneity of the patient populations, but these difficulties are not insurmountable. The N-of-1 design provides a powerful alternative for conventional parallel group randomized controlled trials (RCTs), providing the highest level of evidence. In an N-of-1 study, which is in fact a double-blind RCT requiring less participants, an individual receives multiple blocks of treatment alternated with a comparator (e.g., placebo). This type of research enables investigation of effectiveness of treatments, and addresses the question of inter-individual variability in treatment response.

The ultimate goal is to realize evidence-based care for individuals with RGNDs and ID. This requires a methodological framework for trial designs as well as selection of outcome measures that are (clinically) relevant and measure what matters. This thesis discusses the importance and challenges of evidence-based interventions for RGNDs. **Part I** focusses on N-of-1 studies in RGNDs, starting with providing a methodological framework and recommendations, followed by two study protocols as examples of

conducting a trial within rare and heterogeneous populations. As it is of great importance to measure what really matters in clinical trials, the use of outcome measures for RGNDs and ID is described in **Part II**. Considering that a genetic diagnosis provides opportunities for disorder-specific care, **Part III** explores reporting of genetic diagnoses in multidisciplinary ID care.

Part I: N-of-1 studies

To generate evidence for interventions for RGNDs, this thesis advocates for the use of N-of-1 design. This design is particularly suitable for small patient populations with high variability within and between disorders. It is considered a powerful approach for demonstrating treatment efficacy and can be used to inform personalized clinical treatment decisions. Remarkably, N-of-1 studies have been sporadically reported in RGNDs, as described in **Chapter 2**. To enhance methodological quality, feasibility, personalization and generalizability of N-of-1 studies, recommendations are provided to guide clinicians in realizing evidence-based interventions. For example, we recommend thorough description of baseline characteristics, target symptoms and outcomes, performing a power analysis, consideration of adding a baseline period, dose titration phase, run-in period, washout periods and follow-up measurement. Burden can be minimized by interim analyses and the choice of outcome measures.

As an example of using the N-of-1 design, we started a series of N-of-1 trials with methylphenidate for attention-deficit/hyperactivity disorders (ADHD) in children and adults with Smith-Magenis syndrome (**Chapter 3**). In **Chapter 4**, we provide a protocol for the N-of-1 study with cannabidiol (CBD) for severe behavioral manifestations in children and adults with Tuberous Sclerosis Complex (TSC), mucopolysaccharidosis type III (MPS III), and Fragile X syndrome (FXS). In both double-blind, randomized and placebo-controlled clinical trials, we involved patient (representatives) in trial design and outcome measures.

Although indications and accessibility of the medication differ, both studies aim to provide information on the effectiveness and adverse effects of the intervention. Methylphenidate is first-line pharmacological treatment for idiopathic ADHD. However, efficacy in RGNDs is unclear and there is

increasing evidence for differential treatment response and tolerability. For some practitioners, starting treatment with methylphenidate, with blinded crossover periods, the use of placebo and filling out questionnaires is already part of standard care, and debate is ongoing to what extent an N-of-1 study represents medical research or is part of evidence-based care. By comparison, CBD [Epidyolex®] is currently only approved for refractory epilepsy associated with TSC, Lennox-Gastaut syndrome, and Dravet syndrome. Due to recent market approval, evidence is needed for the effectiveness and adverse effects when used to treat behavioral manifestations. If effective, it is yet unclear whether the medication will be approved and reimbursed for this indication in other individuals with RGNDs and ID.

Part II: Measuring what matters

After identification of treatment targets and drug development, a trial requires appropriately chosen outcome measures to avoid confusion when trial results do not demonstrate benefits, raising the question whether it is due to the inefficacy of the intervention or the chosen outcome measures. For RGNDs and ID, choosing and selecting reliable and valid outcome measurement instruments for trials can be challenging due to the heterogeneity of the populations and cognitive impairments of the individuals.

In **Chapter 5**, an overview of selected outcomes and outcome measurement instruments used in clinical trials in RGNDs, neurometabolic disorders, and ID is provided. Due to the abundancy of instruments and differently used terminology to indicate outcomes, we rather address the overgrowth of outcome measurement instruments to measure the great amount of inconsistently reported outcomes, resulting in a call to action to turn complexity into structure. To illustrate, from 317 clinical trials from the last ten years, we found that 459 different outcome measurement instruments were used of which 290 instruments used in only one clinical trial. Moreover, 438 different outcomes were reported whereas several probably aimed to identify a similar construct. This is problematic for consistency and generalization as well as investigation of measurement properties. Clinical researchers need guidance in choosing appropriate outcome measures

in clinical trials. Recommendations are provided for the near future to select outcomes and instruments to accurately capture what matters to individuals, and an idealized picture is discussed for the long-term.

Existing instruments commonly used in trials in specific RGNDs often focus on single areas, do often not include disease-specific symptoms and may not be sensitive enough for individuals with the particular RGND. Specifically, for the TSC patient organization, researchers, and healthcare providers, the development of an instrument that measures domains and symptoms relevant to individuals with TSC has been considered a top priority. In **Chapter 6**, we describe the development and validation of a disorder-specific patient-reported outcome measure (PROM) for adults with TSC, named the TSC-PROM, that captures the impact of TSC on physical functions, mental functions, activity and participation, the social support someone receives, and health-related quality of life. It consists separate versions in English and Dutch for self-report and proxy-report.

To acquire information from the affected individual's perspective, innovative (user-friendly) ways should be used such as emoticons, symbols or experience sampling, as internal states (e.g., anxiety or depression) may be difficult to measure by proxy-reports. Experience sampling is a method for collecting information about an individual's subjective experiences at multiple (random) times, using digital data collection methods such as smartphones nowadays. **Chapter 7** describes the accessibility and feasibility of experience sampling for the assessment of mental health in individuals with ID, including important gaps in knowledge about acceptability, availability, and appropriateness of current implementation.

Part III: Genetic diagnosis in multidisciplinary ID care

A genetic diagnosis allows disorder-specific care. Due to the rapid technological advances, the genetic etiology can be identified in up to 50% with more than 1500 primary ID genes causing around 1800 ID-related disorders, in addition to another 1250 ID candidate genes, copy number variations, and other genetic causes of ID. Knowing the genetic cause provides information on associated somatic and neuropsychiatric manifestations, providing targets for prognosis, screening, prevention,

monitoring, (targeted) treatment, and care. The often complex and variable multiorgan comorbidity typically requires involvement of a multidisciplinary team including physicians, mental health professionals, and caregivers. **Chapter 8** explores to what extent genetic diagnoses are documented by the care team and identifies associated clinical and demographic factors. It reveals a care gap for personalized care in individuals with ID, showing limited and variable reporting of genetic diagnoses by all types of care providers. We found that more severe levels of ID, lower age, and close proximity of the legal representative's relationship to the client were associated with increased reporting of information about genetic etiology, indicating disparities in access to genetic testing. Early diagnosis is crucial to prevent irreversible damage, especially for metabolic disorders, although relevant for neurodevelopmental disorders as well in terms of monitoring and (preventive) care. We have provided recommendations to overcome barriers and contributors to care gaps to identify individuals at risk of underdiagnosis and undertreatment, and to enable disorder-specific, personalized care and empowerment with regard to diagnostics.

In **Chapter 9**, we discuss the implications of all study findings included in this thesis and provide recommendations and future perspectives, including next steps for the implementation of our findings and a framework to guide clinicians and researchers in future interventional research, such as trial design and outcome measures. We emphasize the need for a specific diagnosis to consider personalized and disorder-specific treatments. To provide evidence-based treatment decisions and to prevent polypharmacy, we advocate the use of N-of-1 studies, considered a much-needed bridge between science and practice, especially in complex patient populations.

This thesis aids in accomplishing clinical research for RGNDs and enabling personalized, disorder-specific care for these vulnerable affected individuals. It discusses challenges involved in evidence-based care for RGNDs. We advocate reporting of genetic diagnosis, including details of diagnostic test results, in individuals with ID, personalized treatment approaches, use of the N-of-1 design, and the thorough selection of appropriate and relevant outcome measures to optimize the potential of personalized medicine for individuals with RGNDs and ID.

Chapter 10

Samenvatting

Wereldwijd zijn er miljoenen mensen met een zeldzame genetische neuro-ontwikkelingsstoornis. Hoewel individueel zeldzaam, komen deze aandoeningen tezamen vaak voor en treffen ze 1-3% van de bevolking. Mensen met deze aandoeningen hebben vaak last van complexe en ernstige comorbiditeiten waarbij meerdere orgaansystemen zijn aangedaan. Over het algemeen veroorzaken de neurologische en psychiatrische manifestaties en de verstandelijke beperking (VB) de grootste last voor de personen met deze aandoeningen en hun naasten. De levenslange behoefte aan vaak intensieve zorg brengt uitdagingen met zich mee voor zorgverleners bij het leveren van optimale, gepersonaliseerde zorg. Bij gepersonaliseerde zorg ligt de nadruk op gezondheidsmonitoring en behandelingen gebaseerd op de oorzaak. Verder onderzoek naar wetenschappelijk onderbouwde zorg is noodzakelijk om de kwaliteit van leven voor deze kwetsbare mensen te verbeteren.

Voor veel zeldzame genetische aandoeningen zijn er steeds meer aangrijpingspunten voor (gerichte) behandelingen. Dit kan variëren van enzym- en gentherapie tot specifieke, vaak strenge, diëten en *drug-repurposing* (het gebruik van bestaande geneesmiddelen voor nieuwe toepassingen). Over de gehele levensloop van iemand met een zeldzame genetische neuro-ontwikkelingsstoornis of VB worden er vaak vele (symptomatische) behandelingen toegepast. Hierbij is er een groot risico op over- en onderbehandeling. Daarnaast is het bewijs voor effectiviteit van ingezette behandelingen vaak beperkt vanwege methodologische en statistische uitdagingen bij deze patiëntpopulaties en lopen mensen mogelijk effectieve (aandoeningsspecifieke) behandelingen mis.

Onderzoek bij zeldzame genetische neuro-ontwikkelingsstoornissen is uitdagend vanwege de kleine patiëntaantallen en de grote onderlinge verschillen in de aan-/afwezigheid en ernst van manifestaties bij mensen met een bepaalde aandoening, maar niet onmogelijk. Het N-*of*-1 design biedt een alternatief voor de grootschalige *randomized controlled trials* (RCT's) en levert in potentie zelfs het hoogste bewijsniveau. In een N-*of*-1 trial, wat in feite een dubbelblinde RCT is met minder deelnemers, ontvangt iemand herhaaldelijk zowel de beoogde behandeling als een niet-werkzame interventie (zoals placebo). Deze onderzoeksmethode gaat in op de vraag van interindividuele variabiliteit in behandelresponse.

Het uiteindelijke doel is om gepersonaliseerde en goed onderbouwde behandelingen te realiseren voor mensen met een zeldzame genetische neuro-ontwikkelingsstoornis of een VB. Dit vereist een methodologisch kader voor enerzijds de onderzoeksmethode (*trial design)* en anderzijds de keuze van relevante uitkomstmaten. In dit proefschrift worden het belang en de uitdagingen besproken van onderzoek naar behandelingen bij deze populatie. **Deel I** richt zich op N-*of*-1 studies bij zeldzame genetische neuro-ontwikkelingsstoornissen, waarbij een methodologisch kader en aanbevelingen worden geboden. Dit wordt gevolgd door twee onderzoeksprotocollen ter illustratie van het uitvoeren van een trial binnen zeldzame en heterogene groepen. Aangezien het van groot belang is om in klinische trials te meten wat er echt toe doet, wordt het gebruik van uitkomstmaten voor zeldzame genetische neuro-ontwikkelingsstoornissen en VB beschreven in **Deel II**. Omdat een genetische diagnose mogelijkheden biedt voor aandoeningsspecifieke zorg, wordt in **Deel III** het rapporteren van genetische diagnoses in multidisciplinaire zorg voor mensen met een VB onderzocht.

Deel I: N-*of*-1 studies

Om bewijs te leveren voor de werkzaamheid van behandelingen bij zeldzame genetische neuro-ontwikkelingsstoornissen wordt in dit proefschrift het gebruik van het N-*of*-1 design besproken. Dit is bij uitstek geschikt voor kleine patiëntpopulaties met hoge variabiliteit binnen en tussen aandoeningen. Het N-*of*-1 design wordt met name beschouwd als een krachtige benadering om de effectiviteit van behandelingen aan te tonen en kan worden gebruikt voor het maken van behandelbeslissingen. N-*of*-1 studies zijn echter sporadisch gerapporteerd bij mensen met zeldzame genetische neuro-ontwikkelingsstoornissen, zoals beschreven in **Hoofdstuk 2**. Om de methodologische kwaliteit, haalbaarheid, personalisatie en generaliseerbaarheid van N-*of*-1 studies te verbeteren, hebben we aanbevelingen opgesteld voor klinisch onderzoekers bij het realiseren van wetenschappelijk onderbouwd interventieonderzoek bij een enkel persoon of patiëntpopulatie. Zo raden wij aan om een gedetailleerde beschrijving te geven van de baselinekarakteristieken, doelsymptomen en uitkomsten, een poweranalyse uit te voeren en een baselineperiode, dosistitratiefase, opbouwperiode, afbouwperiode (wanneer van toepassing), '*washout*´

periode en follow-up periode toe te voegen. De belasting kan worden verminderd door tussentijds analyses te doen om bij genoeg effectiviteit vroegtijdig te kunnen stoppen met de trial, en door geschikte uitkomstmaten te kiezen.

Er is gestart met het opzetten van een reeks N-*of*-1 studies om de effectiviteit van methylfenidaat voor aandachtstekortstoornis met hyperactiviteit (ADHD) te bepalen bij kinderen en volwassenen met het Smith-Magenis syndroom (**Hoofdstuk 3**). In **Hoofdstuk 4** hebben we een protocol opgesteld voor een N-*of*-1 studie naar de effectiviteit van cannabidiol (CBD) voor ernstige gedragsmanifestaties bij kinderen en volwassenen met Tubereuze Sclerose Complex (TSC), mucopolysaccharidose type III (MPS III) en het Fragiele X syndroom (FXS). In beide dubbelblinde, gerandomiseerde en placebo-gecontroleerde klinische onderzoeken zijn patiënten(vertegenwoordigers) betrokken bij het ontwerp van het onderzoek en de keuze van de uitkomstmaten.

Hoewel de indicatie en toegankelijkheid van de medicatie verschillend zijn, hebben beide onderzoeken tot doel informatie te verschaffen over de effectiviteit en bijwerkingen van de behandeling. Methylfenidaat is de eerstelijns farmacologische behandeling voor ADHD, hoewel de werkzaamheid bij zeldzame genetische neuro-ontwikkelingsstoornissen onduidelijk is. Er zijn steeds meer aanwijzingen voor verschillende behandelresponses en verdraagzaamheid op het gebied van bijwerkingen. Voor sommige behandelaren maakt het starten van de behandeling met methylfenidaat, geblindeerde periodes, het gebruik van placebo en het invullen van vragenlijsten al deel uit van de standaardzorg. Het is dan ook de vraag in hoeverre een N-*of*-1 studie (bij geregistreerde middelen in een specifieke populatie) geneesmiddelenonderzoek betreft of dat het deel uitmaakt van op bewijs gebaseerde zorg. Daarentegen is CBD [Epidyolex®] op dit moment enkel geregistreerd voor behandeling van moeilijk behandelbare epilepsie bij TSC, Lennox-Gastaut syndroom en Dravet syndroom. Aangezien CBD recent is toegelaten tot de markt, is bewijs nodig van de effectiviteit en bijwerkingen wanneer het gebruikt wordt om ernstige gedragsproblemen te behandelen. Het is echter onduidelijk of de medicatie goedgekeurd en vergoed zal worden wanneer het voor deze indicatie en bij anderen mensen

met een zeldzame genetische neuro-ontwikkelingsstoornis of VB effectief blijkt te zijn.

Deel II: Meten wat belangrijk is

Bij een veelbelovende behandeling is het kiezen van een geschikte uitkomstmaat cruciaal voor zowel het aantonen van effectiviteit van een behandeling, als aantonen dat een behandeling niet-effectief is. Bij een negatieve onderzoeksbevinding, is de behandeling dan inderdaad niet effectief, of is het een gevolg van verkeerd gekozen uitkomstmaten? Het kiezen van betrouwbare en valide meetinstrumenten kan ingewikkeld zijn bij zeldzame genetische neuro-ontwikkelingsstoornissen of VB vanwege de heterogeniteit van de doelgroepen en cognitieve beperkingen van de mensen.

In **Hoofdstuk 5** wordt een overzicht gegeven van uitkomsten en meetinstrumenten die zijn gebruikt in klinische onderzoeken bij zeldzame genetische neuro-ontwikkelingsstoornissen, neurometabole aandoeningen en VB. De overvloed aan meetinstrumenten en de verschillende terminologieën die worden gebruikt om uitkomsten te duiden, resulteerden in een *call to action*. Van de 317 klinische onderzoeken werden alleen al in de afgelopen tien jaar 459 verschillende meetinstrumenten gebruikt. Hiervan werden 290 meetinstrumenten slechts in één klinisch onderzoek gebruikt. Bovendien werden 438 verschillende uitkomsten gerapporteerd, terwijl sommige waarschijnlijk hetzelfde concept bedoelden. Dit is problematisch voor consistentie, generalisatie en onderzoek naar meeteigenschappen. Klinische onderzoekers kunnen begeleiding nodig hebben bij het kiezen van passende uitkomstmaten. Daarom hebben we aanbevelingen opgesteld voor de nabije toekomst om uitkomsten en meetinstrumenten te selecteren en bespreken we een ideaalbeeld voor de langere termijn.

Bestaande meetinstrumenten die regelmatig worden gebruikt in onderzoeken bij zeldzame genetische neuro-ontwikkelingsstoornissen richten zich vaak op enkele specifieke gebieden, bevatten niet de symptomen die specifiek zijn voor de aandoening en zijn mogelijk niet sensitief genoeg voor mensen met de betreffende aandoening. De TSC-patiëntenorganisatie, onderzoekers en zorgverleners beschouwden de ontwikkeling van een betrouwbaar en

valide meetinstrument dat domeinen en symptomen meet die relevant zijn voor mensen met TSC als topprioriteit. In **Hoofdstuk 6** beschrijven we de ontwikkeling en validatie van deze TSC-specifieke patiënt-gerapporteerde uitkomstmaat (*patient-reported outcome measure;* PROM) voor volwassenen met TSC, genaamd de TSC-PROM. Dit instrument meet de impact van TSC op fysiek functioneren, mentaal functioneren, activiteiten en participatie, sociale steun en gezondheids-gerelateerde kwaliteit van leven. Het bestaat uit afzonderlijke versies in het Engels en Nederlands voor zelfrapportage en rapportage door een proxy.

Om informatie te verkrijgen vanuit het patiëntperspectief zijn innovatieve (én gebruikersvriendelijke) methoden nodig, zoals het gebruik van emoticons en *experience sampling*. Dit is een methode om informatie te verzamelen over de subjectieve ervaringen op meerdere (willekeurige) momenten via bijvoorbeeld digitale gegevensverzamelingsmethoden zoals smartphones. In **Hoofdstuk 7** beschrijven we de toegankelijkheid en haalbaarheid van *experience sampling* voor de beoordeling van de mentale gezondheid bij mensen met een VB, inclusief belangrijke lacunes in kennis over acceptatie, beschikbaarheid en geschiktheid van de huidige implementatie.

Deel III: Genetische diagnose in multidisciplinaire zorg

Een genetische diagnose maakt aandoeningsspecifieke zorg mogelijk. Vanwege de technologische ontwikkelingen kan een genetische oorzaak worden geïdentificeerd bij tot wel 50% van de mensen met een VB. Momenteel zijn er meer dan 1500 primaire VB-gerelateerde genen bekend die tot ongeveer 1800 genetische aandoeningen leiden, naast circa 1250 genen die mogelijk verband houden met een VB, zogenoemde copy number variaties (CNV's), en andere genetische oorzaken van VB. Een genetische diagnose kan informatie bieden over bijbehorende somatische en neuropsychiatrische klachten. Dit helpt bij prognose, screening, preventie, monitoring en (gerichte) behandeling en zorg. Vanwege de vaak complexe en variabele comorbiditeit is betrokkenheid van een multidisciplinair team nodig, waaronder artsen, gedragsdeskundigen, psychologen en begeleiders. In **Hoofdstuk 8** bespreken we in hoeverre genetische diagnoses worden gedocumenteerd door het zorgteam en identificeren we bijbehorende klinische en demografische factoren, waarbij een zorgkloof aangetoond

wordt. Een genetische diagnose was beperkt en niet consistent gerapporteerd door de verschillende soorten zorgprofessionals. Verminderde rapportage van de genetische oorzaak was geassocieerd met een milde VB van cliënten, een hogere leeftijd en geen familielid als wettelijk vertegenwoordiger. Vroege diagnose is cruciaal om onomkeerbare schade te voorkomen. Dit is vooral bij metabole aandoeningen het geval, maar evengoed relevant voor neurologische ontwikkelingsstoornissen op het gebied van monitoring en preventieve zorg. We hebben aanbevelingen opgesteld om barrières weg te nemen die bijdragen aan de zorgkloof. Deze aanbevelingen richten zich er op om risico op onderdiagnose en onderbehandeling te voorkomen en om aandoeningsspecifieke, gepersonaliseerde zorg mogelijk te maken.

In **Hoofdstuk 9** bespreken we de implicaties voor de zorg en onderzoek van alle onderzoeken van dit proefschrift en geven we aanbevelingen en toekomstperspectieven. Vervolgstappen voor de implementatie van onze bevindingen worden besproken, evenals een raamwerk om clinici en onderzoekers te begeleiden bij toekomstig onderzoek naar behandelingen, zoals een studieopzet en uitkomstmaten. We benadrukken de noodzaak van het uitvoeren van genetische diagnostiek zodat gepersonaliseerde en aandoeningsspecifieke behandelingen overwogen kunnen worden. Om onderbouwde behandelbeslissingen te nemen en polyfarmacie te voorkomen, pleiten we voor het gebruik van N-*of*-1-studies. Dit wordt beschouwd als een hoognodige brug tussen wetenschap en praktijk, vooral voor complexe patiëntpopulaties.

Dit proefschrift beoogt bij te dragen aan het mogelijk maken van klinisch onderzoek bij zeldzame genetische neuro-ontwikkelingsstoornissen en gepersonaliseerde, aandoeningsspecifieke zorg. Uitdagingen die gepaard gaan met het onderbouwen van zorg voor deze mensen worden besproken. We benadrukken het belang van (adequate rapportage van) genetische diagnostiek bij mensen met een VB, het gebruik van het N-*of*-1 design en de zorgvuldige keuze van passende en relevante uitkomstmaten. Op deze manier kunnen we gepersonaliseerde zorg voor mensen met zeldzame genetische neuro-ontwikkelingsstoornissen en een VB mogelijk maken.

Addendum

List of contributing authors

Mariëlle Alders
Department of Human Genetics, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands

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

Bas Bijl
Advisium, ‘s Heeren Loo, Amersfoort, The Netherlands

Erik Boot
Advisium, ‘s Heeren Loo, Amersfoort, The Netherlands;
Department of Psychiatry & Neuropsychology, Maastricht University, Maastricht, The Netherlands;
The Dalglish Family 22 q Clinic, Toronto, Ontario, Canada

Marion M.M.G. Brands
Department of Pediatrics, Emma Children’s Hospital, Amsterdam UMC, Amsterdam, The Netherlands;
United for Metabolic Diseases

Hilgo Bruining
Child and Adolescent Psychiatry and Psychosocial Care, Amsterdam UMC location Vrije Universiteit, The Netherlands;
Amsterdam Neuroscience, N=You Neurodevelopmental Precision Center, Amsterdam UMC, Amsterdam, The Netherlands;
Levvel, Center for Child and Adolescent Psychiatry, Amsterdam, The Netherlands;
Emma Center for Personalized Medicine, Amsterdam UMC, Amsterdam, The Netherlands

Jamie K. Capal
Department of Neurology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA;
Department of Neurology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Martina C. Cornel
Department of Human Genetics, Amsterdam UMC location Vrije Universiteit, Amsterdam, The Netherlands

Joost G. Daams
Medical Library, Research Support, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands

Bram Dierckx
Department of Child and Adolescent Psychiatry/Psychology, ENCORE Expertise Center for Neurodevelopmental Disorders, Erasmus Medical Center Sophia Children's Hospital, Rotterdam, The Netherlands

Agnies M. van Eeghen
Advisium, 's Heeren Loo, Amersfoort, The Netherlands;
Department of Pediatrics, Emma Children's Hospital, Amsterdam UMC, Amsterdam, The Netherlands;
Emma Center for Personalized Medicine, Amsterdam UMC, Amsterdam, The Netherlands

Lisa Geertjens
Child and Adolescent Psychiatry and Psychosocial Care, Amsterdam UMC location Vrije Universiteit, The Netherlands;
Amsterdam Neuroscience, N=You Neurodevelopmental Precision Center, Amsterdam UMC, Amsterdam, The Netherlands

Tanjala Gipson
Department of Pediatrics, University of Tennessee Health Sciences Center, Memphis, TN, USA;
Le Bonheur Children's Hospital and Boling Center for Developmental Disabilities, Memphis, TN, USA

Laura C.G. de Graaff
Center for Adults with Rare Genetic Syndromes, Division of Endocrinology, Department of Internal Medicine, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

Mieke M. van Haelst
Department of Human Genetics, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands
Emma Center for Personalized Medicine, Amsterdam UMC, Amsterdam, The Netherlands

Lotte Haverman
Department of Child and Adolescent Psychiatry & Psychosocial Care, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands

Lidewij Henneman
Department of Human Genetics, Amsterdam UMC location Vrije Universiteit, Amsterdam, The Netherlands

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

Bibiche den Hollander
Department of Pediatrics, Emma Children's Hospital, Amsterdam UMC, Amsterdam, The Netherlands;
United for Metabolic Diseases

Leontine W. ten Hoopen
Department of Child and Adolescent Psychiatry/Psychology, ENCORE Expertise Center for Neurodevelopmental Disorders, Erasmus Medical Center Sophia Children's Hospital, Rotterdam, The Netherlands;
Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

Anna C. Jansen
Neurogenetics Research Group, Reproduction Genetics and Regenerative Medicine Research Cluster, Vrije Universiteit Brussel, Brussels, Belgium;
Pediatric Neurology Unit, Department of Pediatrics, Antwerp University Hospital; Translational Neurosciences, University of Antwerp, Antwerp, Belgium

Floor E. Jansen
Department of Pediatric Neurology, Brain, University Medical Center Utrecht, Utrecht, The Netherlands.

Peter Janssens
Department of Nephrology and Arterial Hypertension, Universitair Ziekenhuis Brussel (UZ Brussel), Vrije Universiteit Brussel, Brussels, Belgium

Dick Kampman
Faculty of Science, Operational management, ICT department, Utrecht University, Utrecht, The Netherlands

Clara D.M. van Karnebeek
Department of Pediatrics, Emma Children's Hospital, Amsterdam UMC, Amsterdam, The Netherlands;
Department of Human Genetics, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands;
United for Metabolic Diseases;
Emma Center for Personalized Medicine, Amsterdam UMC, Amsterdam, The Netherlands

Cathelijne Linders
Advisium, 's Heeren Loo, Amersfoort, The Netherlands

Michiel A.J. Luijten
Department of Child and Adolescent Psychiatry & Psychosocial Care, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands

Stijn B. Notermans
Department of Pediatrics, Emma Children's Hospital, Amsterdam UMC, Amsterdam, The Netherlands

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

Wendela L. de Ranitz-Greven
Department of Internal Medicine, University Medical Center Utrecht, Utrecht, The Netherlands

André B. Rietman
Department of Child and Adolescent Psychiatry/Psychology, ENCORE Expertise Center for Neurodevelopmental Disorders, Erasmus Medical Center Sophia Children's Hospital, Rotterdam, The Netherlands

Kit C.B. Roes
Department of Health Evidence, Biostatistics, Radboud University Medical Center, Nijmegen, The Netherlands

Nanda N.J. Rommelse
Department of Psychiatry, Radboud University Medical Centre, Nijmegen, The Netherlands;
Karakter child and adolescent psychiatry, Nijmegen, The Netherlands;

Esther de Rooij-Askes
Advisium, 's Heeren Loo, Amersfoort, The Netherlands

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

Nadia Y. van Silfhout
Department of Child and Adolescent Psychiatry & Psychosocial Care, Amsterdam UMC location University of Amsterdam, Amsterdam, The Netherlands

Caroline B. Terwee
Epidemiology and Data Science, Amsterdam UMC Location Vrije Universiteit, Amsterdam, The Netherlands

Peter M. van de Ven
Department of Data Science and Biostatistics, University Medical Center Utrecht, Utrecht, The Netherlands

Petrus J. de Vries
Centre for Autism Research in Africa (CARA), Division of Child & Adolescent Psychiatry, University of Cape Town, Cape Town, South Africa

Frits A. Wijburg
Department of Pediatrics, Emma Children's Hospital, Amsterdam UMC, Amsterdam, The Netherlands

Marie-Claire Y. de Wit
Department of Pediatric Neurology and ENCORE Expertise Center for Neurodevelopmental Disorders, Erasmus Medical Center Sophia Children's Hospital, Erasmus University Medical Center Rotterdam, Rotterdam, The Netherlands

Janneke R. Zinkstok
Department of Psychiatry, Radboud University Medical Centre, Nijmegen, The Netherlands;
Karakter child and adolescent psychiatry, Nijmegen, The Netherlands;
Department of Psychiatry and Brain Center, University Medical Center Utrecht, Utrecht, The Netherlands

List of publications

This thesis

Müller, A. R., Brands, M. M., van de Ven, P. M., Roes, K. C., Cornel, M. C., van Karnebeek, C. D., ... & van Eeghen, A. M. (2021). Systematic review of N-of-1 studies in rare genetic neurodevelopmental disorders: the power of 1. *Neurology*, *96*(11), 529-540. DOI: 10.1212/WNL.0000000000011597

Müller, A. R., Zinkstok, J. R., Rommelse, N. N. J., van de Ven, P. M., Roes, K. C. B., Wijburg, F. A., ... & van Eeghen, A. M. (2021). Methylphenidate for attention-deficit/hyperactivity disorder in patients with Smith–Magenis syndrome: protocol for a series of N-of-1 trials. *Orphanet Journal of Rare Diseases*, *16*(1), 1-9. DOI: 10.1186/s13023-021-02003-z

Müller, A. R., Luijten, M. A. J., Haverman, L., de Ranitz-Greven, W. L., Janssens, P., Rietman, A. B., … & van Eeghen A.M. (2023). Understanding the impact of tuberous sclerosis complex: development and validation of the TSC-PROM. *BMC Med, 21*, 298. DOI: 10.1186/s12916-023-03012-4.

Müller, A. R., den Hollander, B., van de Ven, P. M., Roes, K. C. B., Geertjens, L., Bruining, H., … & van Eeghen, A. M. (2023). Cannabidiol (Epidyolex®) for severe behavioral manifestations in patients with Tuberous Sclerosis Complex, mucopolysaccharidosis type III and Fragile X syndrome: protocol for a series of randomized, placebo-controlled N-of-1 trials. *Submitted.*

Müller, A. R., van Silfhout, N. Y., den Hollander, B., Kampman, H. C., Bakkum, L., Brands, M. M. M. G., … & van Eeghen, A. M. (2023). Navigating the outcome maze: A scoping review of outcomes and instruments in clinical trials in genetic neurodevelopmental disorders and intellectual disability. *Submitted.*

Bakkum, L., Paalman, C., **Müller, A. R.**, van Eeghen, A. M., & Schuengel, C. (2023). Accessibility and Feasibility of Experience Sampling Methods for Mental Health Research with People with Intellectual Disability: Scoping of Research and Stakeholder Views. *Submitted.*

Müller, A. R., Boot, E., Notermans, S. B., Schuengel, C., Henneman, L., Cornel, M. C., … & van Eeghen, A. M. (2023). Do we care? Reporting of genetic diagnoses in multidisciplinary intellectual disability care: a retrospective chart review. *Submitted*

Other publications

Müller, A., & Van Eeghen, A. (2020). Als alternatief voor grote trials bij mensen met zeldzame genetische syndromen. *TAVG*, *38*(3), 128-131.

Müller, A., Zinkstok, J.R., Rommelse, N.N.J., Van de Ven, P.M., Roes, K.C.B., Wijburg, F.A., Boot, E., & Van Eeghen, A.M. (2020). Methylfenidaat bij kinderen en volwassenen met het Smith Magenis syndroom en ADHD: Een serie van N-of-1 studies. *TAVG, 38*(4), 223-225.

von Scheibler, E. N., van Eeghen, A. M., de Koning, T. J., Kuijf, M. L., Zinkstok, J. R., **Müller, A. R.**, … & Boot, E. (2023). Parkinsonism in Genetic Neurodevelopmental Disorders: A Systematic Review. *Movement Disorders Clinical Practice*, *10*(1), 17-31.

PhD portfolio

Name:	Annelieke R. Müller
PhD period:	September 2019 – September 2023
Promotor:	Prof. dr. F.A. Wijburg
Co-promotores:	Dr. A.M. van Eeghen Dr. E. Boot
Department:	Pediatric Metabolic Diseases

1. PhD training

	Year	Workload (ECTS)
General courses		
Medical Literature (EndNote & PubMed)	2020, 2021	0.2
Basic Course Legislation and Organization (BROK)	2020	1.0
Research Data Management	2020	0.9
Project Management	2020	0.6
The AMC World of Science	2020	0.7
Practical Biostatistics	2020	1.1
Advanced Topics in Biostatistics	2022	2.1
Specific courses		
Peer to peer group coaching	2020	0.5
Amsterdam Public Health: Intersectionality and I	2020	1.0
Infographics	2020	0.1
Children, young patients and family involvement in the design of clinical research	2021	0.3
Observational epidemiology	2022	0.6
N-of-1 for non-statisticians	2022	0.3
Medicinal cannabis	2023	0.3
Seminars, workshops and master classes		
Pharmacogenetics	2019	0.1
Epigenetics	2020	0.1
Human genetics (post-mortem use of genetic and health-related data for research: European policies and perspectives)	2020	0.1
Amsterdam Public Health webinars and workshops, including methodology, reliability studies, precision health, patient involvement, co-creation	2020	0.3
Principles of pediatric clinical pharmacology, NIH	2021	0.1

Fragile X syndrome: physiopathology, therapeutics, innovative therapies and drug candidates	2021	0.5
Academische geneesmiddelenontwikkeling, Nationaal Farmaceutisch Kenniscentrum	2021	3.0
Online presenteren met impact	2021	0.1
Making Science Visual	2021	0.1
Drug repurposing	2021	0.1
CBG-collegedag	2021	0.3
Stofwisselingsziekten, publieksavond	2021	0.1
European Joint Programme on Rare Diseases (EJP-RD) workshops and seminars	2022	0.5
ERN scientific symposium N-of-1 Neurology	2022	0.1
Cannabidiol for Dravet and Lennox-Gastaut syndrome	2022	0.1
Solve-RD: The Treatabolome project; public symposium	2022; 2023	0.3
DORP master protocols	2022	0.1
ERNICA: Clinical trials	2023	0.1
Epicare evidence real-world data	2023	0.1

Presentations

Oral

The power of 1 *Erasmus MC ENCORE, Rotterdam; Radboud UMC, Nijmegen; Amsterdam UMC, Amsterdam*	2020	0.5
The power of 1 *Arts VG regio-overleg, virtual*	2020	0.5
Trial designs voor behandeling van neurologische ontwikkelingsstoornissen *Mini-symposium Alternatieve trial designs voor behandeling van zeldzame ziekten, virtual*	2021	0.5
The power of 1: N-of-1 studies in rare genetic neurodevelopmental disorders & Methylphenidate for ADHD in Smith-Magenis syndrome *International Association for the Scientific Study of Intellectual and Developmental Disabilities (IASSIDD), virtual*	2021	1.0
Het belang van (gepersonaliseerde) uitkomstmaten *Arts VG scholingsdag, Utrecht*	2021	0.5
Cannabidiol for severe behavioral manifestations in RGNDs *Fragiele X expertisenetwerk, virtual*	2021	0.5
Cannabidiol for severe behavioral manifestations in RGNDs *Vereniging Klinische Genetica Nederland, LOG-LOD, virtual*	2021	0.5
Trial designs and outcome measures *22q11 Lage Landen overleg, virtual*	2021	0.5

N-of-1 studies: Methylphenidate for behavioral problems in phenylketonuria *Amsterdam Public Health, webinar*	2022	0.5
TSC-PROM development and validation *TSC-Associated Neuropsychiatric Disorders symposium, virtual*	2022	0.5
Trial designs for rare diseases *European Reference Network (ERN-ITHACA), Budapest, Hungary*	2022	0.5
Implementation of genetic diagnoses in ID care *Amsterdam Kindersymposium, Amsterdam*	2023	0.5
Cannabidiol for complex behavior in FXS *Fragiele X expertisenetwerk, virtual*	2023	0.5
N-of-1 studies: Research or clinical practice? *Amsterdam Public Health Special, Amsterdam*	2023	0.5
De toegevoegde waarde van een genetische diagnose *ZonMW conference 'Zorg voor mensen met een VB', Amersfoort*	2023	0.5
Development and validation of the TSC-PROM *TSC-Associated Neuropsychiatric Disorders symposium, virtual*	2023	0.5
Development and validation of the TSC-PROM *International TSC conference, Washington D.C., USA*	2023	0.5

Poster

The power of 1: N-of-1 studies to improve interventional research for Tuberous Sclerosis Complex *International TSC conference, virtual*	2021	0.5
The power of 1: Systematic review and a study protocol for N-of-1 studies in rare genetic neurodevelopmental disorders *Society for the Study of Behavioural Phenotypes (SSBP) research symposium, virtual*	2021	0.5
Can we make personalized care for individuals with ID happen? Insights from a large ID registry *SSBP research symposium, Oslo, Norway*	2022	0.5
ICT4RARE: Personalized care for individuals with ID *Amsterdam Public Health Annual meeting, Amsterdam*	2022	0.5
Cannabidiol for severe behavioral manifestations in RGNDs; implementation of genetic diagnoses in multidisciplinary ID care (2) *European Reference Network (ERN-ITHACA), Hoofddorp*	2023	0.5
Cannabidiol for severe behavioral manifestations in RGNDs *International TSC conference, Washington D.C., USA*	2023	0.5

(Inter)national conferences

Jonge Onderzoeksdag Erfelijke Aangeboren Aandoeningen, Maastricht	2019	0.3
Amsterdam Kindersymposium, Amsterdam	2020	0.3
Amsterdam Neuroscience Annual Meeting, virtual	2020	0.3

AGEM ImmunoMetNet symposium, virtual	2020	0.3
Amsterdam Kindersymposium, virtual	2021	0.3
Small is beautiful symposium, virtual	2021	1.0
Erfelijke Stofwisselingsziekten Nederland (ESN), voorjaarscongres, virtual	2021	0.3
Congres Goed Gebruik Geneesmiddelen, virtual	2021	0.3
International TSC conference, virtual	2021	1.0
International Association for the Scientific Study of Intellectual and Developmental Disabilities (IASSIDD), virtual	2021	1.0
Society for the Study of Behavioural Phenotypes (SSBP) research symposium, virtual	2021	1.0
Jubileumsymposium NVK-sectie Erfelijke en Aangeboren Aandoeningen, Nijmegen	2022	0.5
TSC-Associated Neuropsychiatric Disorders symposium, virtual	2022	0.1
Jonge Onderzoekersdag, Rotterdam	2022	0.3
SSBP research symposium, Oslo, Norway	2022	1.0
Symposium Medicijn voor de Maatschappij, Amsterdam	2022	0.1
Amsterdam Kindersymposium, Amsterdam	2023	0.3
Symposium ZeldSamen, Amersfoort	2023	0.3
Dutch Neurodevelopmental Day, Amsterdam	2023	0.1
Congres Goed Gebruik Geneesmiddelen, Den Bosch	2023	0.3
ERN-ITHACA workshop, Hoofddorp	2023	0.5
N=1 symposium, virtual	2023	1.0
Emma Center for Personalized Medicine symposium, Amsterdam	2023	0.1
ZonMW congres 'Zorg voor mensen met een VB', Amersfoort	2023	0.3
TSC-Associated Neuropsychiatric Disorders symposium, virtual	2023	0.1
International TSC conference, Washington D.C. USA	2023	1.0
SSBP research symposium, virtual	2023	1.0

2. Teaching

	Year	ECTS
Lecturing		
Arts VG scholingsdag	2021	0.5
Lecture master programme Neuroscience	2022	0.5
Supervising		
Michiel Hölscher, masterthesis	2022	1.0
Stijn Notermans, masterthesis	2022	1.0
Maurits Endlich, masterthesis	2023	1.0
Brigit Thomassen, masterthesis	2023	1.0

3. Parameters of Esteem

	Year
Grants	
Co-applicant: Methylphenidate for ADHD symptoms in phenylketonuria, Amsterdam Public Health, PI A. van Eeghen	2021
Co-applicant: OUTCOMES4RARE, For Wishdom foundation, PI A. van Eeghen	2021
Co-applicant: CBD-trial, Stichting TSC Fonds, PI A. van Eeghen	2021
Co-applicant: GAS4VB, Cornelia-Stichting, PI A. van Eeghen	2022
Travel grant: Amsterdam Public Health Personalized Medicine	2023

4. Publications

	Year
Peer reviewed	
European Journal of Paediatric Neurology	2021
Orphanet Journal of Rare Diseases	2022
Other	
Gastredacteur Tijdschrift voor Artsen Verstandelijk Gehandicapten (TAVG)	2020
Training Upcoming Leaders In Paediatric Science (TULIPS) PhD curriculum	2021-2023

Research data management

The research followed ethical guidelines and was conducted in accordance with the principles of the Declaration of Helsinki. Research data management was conducted according to the Findability, Accessibility, Interoperability, and Reusability (FAIR) principles when applicable and possible. All manuscripts have been submitted open access, and open access supplementary materials were provided for transparency and reusability. Data sheets are available upon request from the corresponding author. For the TSC-PROM study and clinical trials, LimeSurvey, Research Electronic Data Capture (REDCap), and Castor EDC were used, respectively, which are secure, web-based software platforms designed to support data capture for research.

Figures were created with Draw.io and Flaticon.com.

Curriculum Vitae

Annelieke Rosalie Müller was born on February 27, 1994 in Amsterdam, the Netherlands. Together with her brother Jasper, she was raised in Purmerend. From the age of four, she started playing music and in 2004, she was admitted to the Royal Conservatory of The Hague to study violin with Coosje Wijzenbeek. In 2007, she studied at the young talent division, Sweelinck Academy at the Conservatory of Amsterdam. Thanks to the endless support and enthusiasm of her parents, she was able to perform several times a month in both the Netherlands and abroad, and won prices at (inter)national music competitions.

In 2012, she obtained her Atheneum degree at the Da Vinci College Purmerend and started Psychobiology at the University of Amsterdam. During her bachelor's programme, she was member of the student council of the Faculty of Science. She did an internship at the University Medical Center (UMC) Utrecht and studied the effectiveness of the brain stimulation 'transcranial direct current stimulation' (tDCS) on auditory hallucinations in psychotic disorders. After receiving her bachelor's degree, she worked as a research and teaching assistant and was a board member (treasurer) at the ASVA student union. She started her research master Brain and Cognitive Sciences at the Institute for Interdisciplinary Studies at the University of Amsterdam. In her first research project at the Netherlands Institute for Neuroscience, she studied hypothalamic glial expression in mood disorders and suicide, under supervision of prof. dr. Dick Swaab. Her second research project was about presynaptic proteins and inhibitory pathways in hippocampal sclerosis at the University of British Columbia, Vancouver, Canada, supervised by prof. dr. William Honer. Furthermore, she has worked at the autopsy team of the Netherlands Brain Bank for six years.

After graduating cum laude in 2019, she started her PhD at 's Heeren Loo and the department of pediatric metabolic diseases, supervised by prof. dr. Frits Wijburg, dr. Agnies van Eeghen, and dr. Erik Boot, which has led to this thesis. During that time, she studied trial methodology to foster interventional research for rare genetic neurodevelopmental disorders. To increase her knowledge on health research methodology, she followed several courses at the Graduate School of the Amsterdam UMC. She acquired a travel grant and other research grants (PI dr. Agnies van Eeghen) to enable several research projects for genetic neurodevelopmental disorders. She presented her work (inter)nationally, such as at the Society for the Study of Behavioural Phenotypes (SSBP) in Oslo (Norway) and virtually due to Covid-19, the International Association for the Scientific Study of Intellectual and Developmental Disabilities (IASSIDD), and the International Tuberous Sclerosis Complex conference in Washington D.C. (USA). She was also invited to present on trial designs for rare disorders at the European Reference Network (ERN-ITHACA) board meeting in Budapest, Hungary.

In her free time, Annelieke enjoys playing violin, sports (climbing, cycling, swimming, and running), being around with friends, and travel to the mountains. Since 2021, she lives in Amersfoort together with Dick Kampman, who is a software developer at the Utrecht University.

Dankwoord

Dit proefschrift was niet mogelijk geweest zonder betrokkenheid van een aantal personen. Daarom wil ik de laatste pagina's benutten om een aantal bijzondere contacten en samenwerkingen aan te halen en belangrijke personen te bedanken.

Allereerst **prof. dr. Wijburg**, beste Frits, vanaf het moment dat u mijn promotor werd, hebt u mij wegwijs gemaakt in de verschillende werelden van het ziekenhuis. De overleggen waarbij we naast het bespreken van de voortgang ook konden uitzoomen, waren waardevol voor mij. Uw intelligentie, integriteit en betrokkenheid zorgden voor wijze adviezen, keuzes en beslissingen. En sorry dat wij vaak voor (open access) journals kozen waar het Amsterdam UMC (nog) geen overeenkomst mee had, maar heel fijn dat u dit mogelijk wilde maken. Ik wil u hartelijk danken voor alle begeleiding en wijsheid. Ik ben vereerd dat ik uw laatste promovenda mocht zijn en ik wens u een mooi en welverdiend pensioen toe.

Daarnaast wil ik mijn twee copromotors, **dr. Agnies van Eeghen** en **dr. Erik Boot**, enorm bedanken voor de mogelijkheid om het promotieonderzoek bij jullie te doen. Ondanks dat de vacature bij 's Heeren Loo in principe voor een arts-onderzoeker was, hebben jullie mij het vertrouwen gegeven om ook als neurobioloog aan dit belangrijke onderwerp te werken. Jullie hebben mij intellectueel uitgedaagd en ik heb het erg op prijs gesteld om met jullie samen te mogen werken. Jullie zijn inhoudelijk enorm sterke, begripvolle, respectvolle en tegelijkertijd ook gezellige collega's. Ik denk dat het vooral aan jullie te danken is geweest dat artikelen over het algemeen vrij makkelijk geaccepteerd werden. En het olijfboompje dat in coronatijd bezorgd werd voor mijn eerste publicatie zal ik ook niet snel vergeten.
Agnies, ik vond het fantastisch om samen met jou te kunnen pionieren. Onder jouw begeleiding is mijn promotieavontuur begonnen. Ik heb in de jaren die volgden ontzettend veel geleerd en dat was zonder jouw begeleiding niet mogelijk geweest. Ik wil je enorm bedanken voor de uitdagingen, leuke en leerzame discussies, je relativisme en humor, de vrijheid die je mij hebt gegeven en de kansen die je hebt geboden. Dat heb ik altijd erg gewaardeerd. Het is bewonderingswaardig hoe jij alle ballen in de lucht weet te houden en

altijd weer met nieuwe ideeën en zeer belangrijke projectvoorstellen komt om de zorg voor mensen met een verstandelijke beperking te verbeteren. Als ik het proces traag vond gaan, bleef jij stimuleren met 'stapje voor stapje'. We zijn in de afgelopen jaren steeds dichter naar elkaar toe gegroeid en een (h)echt team geworden. Ik voel me vereerd dat ik jouw eerste promovenda mocht zijn. Jij hebt mij 'grootgebracht' en ik hoop dat we nog lang kunnen blijven samenwerken. Jouw enthousiasme is aanstekelijk.
Erik, jouw kritische blik en gevatte opmerkingen hebben het onderzoek telkens naar een hoger niveau gebracht. Met jouw kennis en expertise zijn de artikelen sterk verbeterd. Je hebt geholpen het doel scherp voor ogen te houden en de projecten reëel te houden en af te kaderen. Het gaf me altijd veel vertrouwen als jij akkoord was met een artikel, want dan wist ik dat er écht geen fouten meer in stonden en het inhoudelijk klopte. Je bent een betrokken, fijne en betrouwbare collega en ik wil jou enorm bedanken voor de mooie vier jaar samenwerking.

Beste leden van mijn promotiecommissie, **prof. dr. Hilgo Bruining, prof. dr. Lidewij Henneman, prof. dr. Jaap Groothoff, prof. dr. Nicole Wolf, prof. dr. Martin Offringa, dr. Sylvia Huisman,** hartelijk dank voor het lezen en beoordelen van mijn proefschrift. Het is een eer om dit proefschrift tegenover u te mogen verdedigen. Prof. dr. Martin Offringa, veel dank dat u vanuit Canada naar Nederland wilde reizen om aanwezig te zijn bij de promotieplechtigheid.

Veel dank aan **'s Heeren Loo** die dit promotieonderzoek mogelijk heeft gemaakt. De mogelijkheid om vanuit een zorginstelling innovatief wetenschappelijk onderzoek te doen, levert een unieke bijdrage aan zowel de wetenschap als de zorg voor mensen met een verstandelijke beperking. Het maatwerk dat 's Heeren Loo aan de cliënten levert, zag ik ook terug in het maatwerk dat aan de ondersteuning van onderzoekers werd gegeven. Ook gaat mijn dank uit naar de wetenschappelijke adviesraad voor de kritische vragen op het onderzoek gedurende het traject.
Het promotieonderzoek vanuit 's Heeren Loo begon bij de opzetting van het expertisecentrum en polikliniek Genetische syndromen door **dr. Erik Boot** en **dr. Agnies van Eeghen** (de "papa en mama" van de poli), met **dr. Claudia Vingerhoets, Zinzi Vink** en **Emma von Scheibler** (eerste-generatie-

promovenda bij de poli Genetische syndromen van 's Heeren Loo). Al snel werd dit kleine team uitgebreid tot een ware multidisciplinaire poli. **Egbert Broers**, **Gerdie van Achterberg – Blom**, **Esther de Rooij-Askes**, **Barber Tinselboer**, **Jiske van der Meulen**, **Mieke Veltmeijer**, **Reggy Gargosky**, **Cathelijne Linders**, **Sandra Kruithof**, **Anneke Janssen**, **Judith Soomer**, **Hester Jaspers Faijer-Westerink**, **Malu van Schaijk**, **Hadassa Kwetsie**, **Bojana Milojkovic-Kerklaan**, **Nelly Oorbeek**, **Violetta van Wichen**, **Heidy Buitenhuis**, **Edith Rijnsburger** en natuurlijk **Bas Bijl**, ik wil jullie hartelijk bedanken voor de betrokkenheid. Het was een plezier om met zulke gedreven en fijne collega's te mogen samenwerken.

Beste patiëntorganisaties, in het bijzonder **Stichting Smith Magenis Syndroom Nederland**, **Stichting Tubereuze Sclerosis Nederland** (STSN) en **Fragiele X Vereniging Nederland**, heel veel dank voor het meedenken, de input op de onderzoeken, de financiële middelen en de hulp bij de werving.

Beste co-auteurs, dear co-authors, thank you for the inspiring and valuable teamwork and discussions to bring the research projects to a higher level.

Beste **N-of-1 expert board**, beste **prof. dr. Clara van Karnebeek**, **prof. dr. Martina Cornel, prof. dr. Kit Roes, prof. dr. Frits Wijburg**, **prof. dr. Dirk Lefeber**, **dr. Agnies van Eeghen**, **dr. Marion Brands**, **dr. Peter van de Ven**, **dr. Charlotte Ockeloen**, **Vincent van der Wel** en **Bibiche den Hollander**, ik wil jullie enorm bedanken voor het delen van alle expertises. Ik heb veel geleerd van de interessante discussies die we de afgelopen jaren hebben gevoerd. Het begon met de systematische review en verkenning van N-of-1 trials bij zeldzame genetische neuro-ontwikkelingsstoornissen en leidde tot vele nieuwe projecten en verschillende trials.

Beste **MPHSMS-team**, hartelijk dank voor jullie betrokkenheid en samenwerking aan de N-of-1 studie met methylfenidaat voor kinderen en volwassenen met het Smith-Magenis syndroom (SMS). **Esther de Rooij-Askes**, jij kwam als eerste met een concreet voorstel voor een N-of-1 trial. Veel dank voor je vertrouwen om dit uit te gaan voeren, ik vond het fijn om dit samen met jou te starten en de samenwerking die volgde. Veel dank voor jouw kennis en expertise met SMS. Beste **prof. dr. Nanda Lambregts-Rommelse**,

dr. **Janneke Zinkstok**, **Cathelijne Linders, Zinzi Vink** en **Reggy Gargosky**, ik wil jullie hartelijk danken voor de prettige samenwerking en jullie kennis en ervaring op het gebied van de problematiek bij deze kinderen en volwassenen. Een N-of-1 trial vergt soms veel afstemming en contact, maar ik ben blij dat we dit mogelijk hebben kunnen maken en we duidelijkheid hebben kunnen geven aan de deelnemers over de effectiviteit van methylfenidaat bij hen.

Beste **CBD4RARE-team**, hartelijk dank voor de samenwerking en het delen van jullie expertises. In het bijzonder dank aan **Bibiche den Hollander, dr. Marion Brands, dr. Floor Jansen en dr. Marie-Claire de Wit** voor de inhoudelijke discussies en bijdragen. We hebben enig geduld moeten hebben, maar hopelijk kan het project snel van start gaan. Ook dank aan **Jazz Pharmaceuticals** en **Stichting TSC Fonds** voor het mogelijk maken van dit onderzoek.

Dear **collaborators of the TSC-PROM**, many thanks to all your contributions to this project initiated by **dr. Agnies van Eeghen**. **Stichting TSC Fonds**, thank you for financially enabling this project. **Dr. Lotte Haverman** and **dr. Michiel Luijten**, without your assistance, the validation would not have been possible. Special thanks to **dr. Wendela de Ranitz-Greven**, **dr. Tanjala Gipson**, **dr. Jamie Capal**, **Molly Griffith**, **Lauren Davis**, and **Jo Anne Nakagawa** (**TSC Alliance**), **Eva Schoeters** (**Be-TSC**), and **Jan-Paul Wagenaar** (**STSN**) on behalf of the patient organizations for (international) recruitment. En natuurlijk, **Marloes Reurs**, heel veel dank voor de fijne overdracht van dit project.

Daarnaast wil ik eenieder bedanken die heeft bijgedragen aan één van de projecten. **Prof. dr. Mieke van Haelst** en **dr. Mariëlle Alders**, veel dank voor jullie inhoudelijke bijdragen op het gebied van de klinisch genetica. **Prof. dr. Carlo Schuengel** en **dr. Lianne Bakkum**, hartelijk dank voor de fijne samenwerkingen op het gebied van uitkomstmaten, experience sampling methods en rapportage van genetische diagnostiek bij de populatie met een verstandelijke beperking. **Prof. dr. Hanneke van der Lee, prof. dr. Annemieke Buizer** en **dr. Charlotte Gaasterland**, heel veel dank voor de inspiratie en voorzet om Goal Attainment Scaling te gaan implementeren voor onderzoek bij zeldzame aandoeningen. **Dr. Leonie Menke** en **dr. Sylvia**

Huisman, heel veel dank voor de fijne samenwerkingen en interessante discussies. Leonie, dankjewel dat ik jouw werkplek warm mocht houden tijdens de eerste maanden van mijn promotieonderzoek toen jij in Nieuw-Zeeland zat. **Dr. Leontine ten Hoopen, dr. André Rietman, dr. Bram Dierckx, dr. Laura de Graaff, dr. Marie-Claire de Wit**, veel dank voor de prettige samenwerkingen op het gebied van TSC en FXS.
Ook veel dank aan de verschillende samenwerkingspartners zonder wie de projecten niet mogelijk waren geweest. **Prof. dr. Carla Hollak**, de trial apotheek van het Amsterdam UMC, het Kenniscentrum Geneesmiddelenonderzoek, Medicijn voor de Maatschappij, Clinical Cannabis Care, de Transvaalapotheek, Tiofarma, legal research support, het onderzoeksbureau VKC en in het bijzonder **Marjolein Spiering**, heel veel dank voor de samenwerkingen en bijdragen!

Veel dank ook aan **TULIPS** voor het inspirerende PhD curriculum. Lieve Tulipers, **Annelotte, Hans Jacob, Imke, Inge, Konradin, Lauren, Lien, Mala, Martijn, Pauline, Romy, Serife, Trude en Vera**, wat een fijne groep zijn jullie. Ik keek erg uit naar onze bijeenkomsten door het land. Hopelijk kunnen we dit voortzetten in Noorwegen en Engeland en blijven we elkaar tegenkomen binnen de child health.

Lieve mede-promovendi, metabollies en H8 kamergenoten, **Lotte, Hadassa, Nadia, Bibiche, Mirthe, Imeze, Merel, Doortje, Berith, Heleen, Sibbeliene, Charlotte, Elise, Laura, Liz, Roxanne, Sanne, Adinda, Cunera, Lieke, Said, Anne, Sophie, Claire, Daniël, Menne, Marsh, Fenne, Kelly, Daniëlle, Laurens, Paul en Marie-Louise,** dank voor de afgelopen jaren, gezelligheid en samenwerkingen! Het begon bij de luxe lunches en werd al snel gevolgd door de borrels, salsadansen, dinertjes, cocktails maken en de babyshower. Ik heb ook genoten van de samenwerkingen die binnen de kindermetabole/EAA-groep zijn ontstaan, waarbij ik bijvoorbeeld kon sparren met Hadassa als het ging over veroudering, met Nadia over uitkomstmaten, met Bibiche over N-of-1 studies en Mirthe over richtlijnen en evidence-based medicine. Ik ben blij dat ik de laatste wetenschappelijke tastings (welke eierkoeken, cola en chips zijn het lekkerst?) nog heb mee kunnen doen en kijk uit naar het uiteindelijke rapport. Ik ga onze cappuciens bij Albert(ino?) missen, Sibbeliene. Veel succes met jullie promotieonderzoeken.

Lieve vrienden, dank voor alle gezelligheid en steun de afgelopen jaren. **Yara, Nicole, Saartje, Desirée, Anouk, Martha, Naomi, Sophia**, wat prijs ik me gelukkig om jullie als hersenvriendjes te hebben. Ook al wonen we grotendeels inmiddels overal behalve in Amsterdam, het is altijd zo fijn om met jullie te zijn. Om jou, Nic, als klimmaatje te hebben is heel leuk! Onze sessies zijn altijd een leuke afleiding waarin we ook verder kunnen dromen over mogelijke samenwerkingen, dus snel herstellen zodat we naar de 6C+ kunnen werken. **Margo, Lydia, Jeske, Karlijn, Floor en Djamila**, dezelfde studie, dezelfde interesses en bijna dezelfde promotiedatums, het was leuk om de PhD-avonturen met jullie te kunnen delen. En **Marjon**, bedankt voor jouw betrokkenheid en wijze adviezen. Ik ben blij dat je zo een fijne leidinggevende voor Dick (en mij soms ook een beetje) bent. **Lisa en Tessa**, wat leuk dat wij nog steeds contact hebben gehouden sinds de congrescommissie jaren geleden, en dat jullie ook in het onderzoek zitten! Lieve **Henriëtte** en **Jim**, de leuke en muzikale klimdates met jullie zijn onvergetelijk, waarin we de tijd helemaal kunnen verliezen en tot het uiterste weten te gaan. **Mariëlle** en **Klaasjan**, wat was het leuk om elkaar vanuit de LSVb te kunnen blijven zien in 033. Ook veel dank aan je lieve ouders voor de wijze adviezen, Mariëlle! Lieve **Bianca, Linda, Marieke** en **Janneke**, een goede basis is het halve werk! Het is fijn om jullie al bijna 16 jaar als vriendinnen te hebben, dat zit dus wel snor! Lieve **Lisanne** en **Stephano**, al sinds mijn 12e ben je een heel goede vriendin en wat hebben we een lol gehad. Op naar nog 18 jaar vriendschap!

Lieve **Lotte, Yara** en **Bibiche**. Ik ben heel blij dat jullie mijn paranimfen willen zijn. Lotte, jij was mijn eerste maatje op het AMC. Ik kan altijd alles bij jou kwijt en jouw openheid en betrokkenheid maken jou een heel fijne collega. Yaar, jouw oneindige interesse in anderen en enthousiasme om werkelijk álles zijn aanstekelijk. Je staat altijd voor een ander klaar. Zo ook toen ik een nieuwe analyse wilde uitvoeren in R en jij daar een hele dag voor vrij hebt gemaakt, dankjewel! Bibiche, N-of-1 maatje, ik vond het heel fijn om samen met jou aan N-of-1 studies te werken en de CBD-trial te trekken. Ook al hebben we nog niet de trial uitgevoerd, de fundering ligt er. Dankjewel ook voor al jouw goede stickerbijdragen – die heb ik stuk voor stuk overgenomen.

Natuurlijk ook veel dank aan **Erik**, **Ellen**, **Margot**, **Hugo** en **Kelly** voor jullie steun en gezelligheid in Amersfoort met de veelvoudige kopjes koffie en wijntjes. Door jullie enthousiasme werd ik uitgedaagd om mijn verhaal kort en krachtig uit te leggen, waarbij ik mijn elevator pitch goed kon oefenen. Daar heb ik tijdens mijn promotieonderzoek veel profijt van gehad. De gemoedelijke sfeer en nuchtere kijk hebben mij ook geholpen te ontspannen en relativeren. Ik kom graag weer langs op Terschelling (en dan zonder publicatiedeadline).

Lieve **Jasper** en **Berber**, heel erg bedankt voor jullie betrokkenheid en steun. Dankjewel voor je luisterend oor en goede adviezen waarmee je me vaak aan het denken weet te zetten, Ber. Lieve broer, ik wil je enorm bedanken voor jouw bijdragen. Ik denk dat je niemand hoeft te overtuigen van jouw kwaliteiten, het visueel overbrengen van je boodschap en ontwerpen (zie kaft) naast de huisvestingsprojecten en grote verbouwingen op het AMC die jij allemaal in goede banen weet te leiden. Ook heel leuk dat jij je tekenskills hebt kunnen overdragen bij de workshop aan ons team van 's Heeren Loo. En wat is het leuk dat jij ook op het AMC werkt en we daardoor zoveel koffiedates ☕ hebben kunnen doen.

Lieve **Frans** en **Ellen**, heel veel dank voor al jullie onvoorwaardelijke vertrouwen, betrokkenheid, ondersteuning, optimisme, relativisme en liefde. Jullie zijn altijd begaan met Jasper en mij en hebben ons gestimuleerd om te doen wat we willen en het mogelijk gemaakt om die kansen aan te grijpen. Toen ik begon aan mijn promotieonderzoek hebben jullie direct een schema gemaakt van de projecten waarmee ik bezig was en hoe die zich tot elkaar verhouden. Als een artikel gepubliceerd werd, wilden jullie het meteen lezen, zelfs het artikel over de validatie van een uitkomstmaat. Als ik weer een keer te optimistisch had gepland, wist ik dat ik op jullie kon rekenen (zoals bijvoorbeeld last-minute een wetenschappelijke poster ophalen voor een congres, maar ook natuurlijk de duizenden ritjes van Purmerend naar Hilversum en elders in het land voor de vioollessen vroeger). Het is bewonderingswaardig hoe jullie altijd het beste uit het leven halen en hoe jullie met elkaar en anderen omgaan. Na ruim 50 jaar samenzijn is het ook niet gek dat jullie afzonderlijk van elkaar toch altijd hetzelfde weten te appen. Bij jullie is het glas altijd (half)vol. Jullie zijn een voorbeeld voor mij en vele anderen.

En laatste, maar zeker niet de minste, lieve **Dick**, ik mag geen 'dank' van je zeggen, maar ik wil je toch heel graag laten weten dat ik enorm veel aan jou heb gehad, hoe blij ik met je ben en hoe belangrijk jij voor mij bent. We leerden elkaar net kennen voordat ik klaar was met mijn master en ik voor de 'beroepskeuze' stond, waarin jij mij het vertrouwen gaf en hielp bij de keuze. Al snel kwam covid, maar misschien pakte dat voor ons juist wel goed uit. Je hebt mij uit 020 gekregen en ik vind het heel fijn om samen met jou in Amersfoort te wonen. Je hebt mij altijd mentale support gegeven en liet me vrij om ook in de avonden of weekenden te werken. Tegelijkertijd wist je ook balans aan te brengen wanneer dat nodig was. Of wanneer we juist wel of niet wat konden 'combineren', zoals de vakantie en het congres in Noorwegen. Je hebt heel veel humor, een eindeloos geduld (als ik toch nog even wat af wilde maken), maar je bent ook relativerend en meedenkend. De scoping review bijvoorbeeld. Jouw hulp heeft ons gered en de dataverwerking weer mogelijk gemaakt (en ja, de weddenschap resulteerde er in dat ik nu ook echt meer moet gaan programmeren). Hoe leuk ook dat wij co-auteurs zijn en dit jouw eerste artikel wordt! En altijd als er iets met de computer of software was, wist jij het op te lossen of kwam jij met een slim idee. Onze wandelingen om uit te zoomen en alles op een plekje te krijgen betekenen veel voor mij. En nu ben jij ook helemaal thuis in het jargon en de acroniemen die wij voor de onderzoeken gebruiken. Je weet waarschijnlijk meer van fenylketonurie en tubereuze sclerose complex dan mening ander. Lieve Dick, zonder jou was dit echt niet mogelijk geweest, ik hou van je, heel veel DANK!

Lieve allemaal, ik wil jullie enorm bedanken voor alles.

ISBN/EAN: 978-90-833743-7-6

's Heeren Loo