DIAMOND

Development of generalisable methodology for n-of-1 trials delivery for very low volume treatments

Development of a generalisable methodology for n-of-1 trials delivery for very low volume treatments (DIAMOND)

Authorship:

The research was conducted by:

Liv Hawksworth – Research Assistant, Sheffield Clinical Trials Research Unit, The University of Sheffield.

Professor Steven Julious – Professor of Medical Statistics, School of Health and Related Research, The University of Sheffield.

Robin Chatters – Research Fellow, Sheffield Clinical Trials Research Unit, The University of Sheffield.

Professor Stephen Senn – Honorary Professor, School of Health and Related Research, The University of Sheffield.

Dr Andrew Cook – Associate Director, Southampton Clinical Trials Unit.

Abstract

Introduction

n-of-1 trials are multi-period crossover trials designed to evaluate health technologies within individual patients. The DIAMOND project (**D**evelopment of a general**is**able **m**eth**o**dology for **n**-of-1 trials **d**elivery for very low volume treatments) was undertaken to facilitate the application of n-of-1 trials, particularly in rare disease research, by producing guidance on the design, implementation, and analysis of n-of-1 trials.

Methods

We conducted a review of the characteristics of randomised n-of-1 trials published between January 2011 and May 2021. Following the review, we developed a set of key points to consider to guide the design and implementation of n-of-1 trials. These key points were informed by a stakeholder workshop, collaborator discussion, and a stakeholder dissemination and feedback event. The stakeholder workshop sought to gain the perspectives of a variety of stakeholders (including clinicians, researchers, and patient representatives) on the design and use of n-of-1 trials. A discussion between the study team and collaborators was held to reflect on the workshop. Lastly, the stakeholders from the workshop were invited to a dissemination and feedback session. During this session, the proposed key points were presented to stakeholders and their feedback was requested. The feedback received was incorporated into the final set of key points.

Results

The review identified 52 randomised n-of-1 studies, 8 of which (15.4%) were conducted in rare diseases. There were no apparent differences in trial design according to whether they were conducted in a rare disease or not. Workshop discussions centred on the types of questions that n-of-1 trials can be used to answer, the treatments they can be used to assess, and potential outcomes of n-of-1 trials.

A set of key points were developed based on the results of the review and insights from the workshop and subsequent discussions. They provide guidance on when an n-of-1 trial might be a viable or appropriate study design and discuss key decisions involved in the design of n-of-1 trials, including determining an appropriate number of treatment periods and cycles, the choice of comparator, recommended approaches to randomisation and blinding, the use of washout periods, and approaches to analysis.

Additional outputs of the project include two courses on n-of-1 trials and materials to support the analysis of n-of-1 studies in SAS, Genstat and R.

Discussion

The key points developed in the project will support clinical researchers to understand key considerations when designing n-of-1 trials. It is hoped they will support the wider implementation of the study design.

Table of Contents

List of Abbreviations

NIHR – National Institute for Health and Care Research

ISRCTN – International Standard Randomised Controlled Trial Number

PPI – Patient and Public Involvement

RCT – Randomised Controlled Trial

CONSORT – Consolidated Standards of Reporting Trials

CENT – Consort Extension for n-of-1 Trials

BLUP – Best Linear Unbiased Prediction

Plain English Summary

n-of-1 trials test the effects of a treatment by following a single patient over time. The patient will receive the treatment that is being tested and a comparator treatment (e.g. placebo) multiple times. The order in which the treatments are given is decided by chance. Doctors and researchers are sometimes unsure when and how to use n-of-1 trials. We therefore undertook the DIAMOND (**D**evelopment of a general**isa**ble **m**eth**o**dology for **n**-of-1 trials **d**elivery for very low volume treatments) project to develop guidance for doctors and researchers.

We started by finding n-of-1 trials that have been planned or carried out since 2011 in order to understand their features and how they were designed. We shared our findings at a workshop with doctors, researchers and patient representatives who helped us think about some key points that would help other doctors and researchers learn about these trials. We drafted the list of key points then asked the doctors, researchers and patients for feedback.

The final set of key points provides guidance as to when an n-of-1 trial is appropriate, that questions that can be addressed, and how the trial may be designed and interpreted. We have also developed other resources for researchers, including detailed statistical guidance on how to analyse n-of-1 trials, and statistical software packages, and training sessions for both statisticians and non-statisticians.

Scientific Summary

Background

Informal evaluations of the effects of treatment on a patient are common in routine clinical practice (1). However, these 'trials of therapy' are susceptible to bias and consequently have limited use in determining the optimal treatment for a particular patient (1).

n-of-1 trials combine research processes (e.g., blinding and randomisation) with the investigation of treatment effects in an individual patient to reduce bias and consequently improve clinical decision making. They have a within-patient, multi-period, crossover design in which a single patient switches between different treatment conditions until an individual treatment effect is established. The treatment allocation to study periods is usually randomised; these n-of-1 trials can be described as randomised controlled trials (RCTs) conducted in an individual patient, where the patient acts as their own control.

n-of-1 trials have the potential to evaluate health technologies in individual patients and therefore increase therapeutic precision (2,3). They may be particularly useful in rare disease research, in which there are limited options in terms of study design due to small patient populations (4).

It has been highlighted that n-of-1 trials are an underused study design (5). Barriers to the implementation of n-of-1 trials might be behind their underuse, such as clinicians not believing in the clinical value of n-of-1 trials, and a lack of understanding of the technical and statistical elements of the design (6). These barriers might be amenable to influence through guidance on the design and use of n-of-1 trials.

The DIAMOND project (**D**evelopment of a general**isa**ble **m**eth**o**dology for **n**-of-1 trials **d**elivery for very low volume treatments) was undertaken to facilitate the application of n-of-1 trials, particularly in rare disease research, by producing guidance on the design, implementation, and analysis of n-of-1 trials.

Objectives

The objectives were to:

a) Review previously conducted n-of-1 trials to establish their characteristics;

b) Conduct a series of stakeholder workshops to develop a list of key points to consider for n-of-1 trials;

c) Provide training courses to disseminate the findings;

d) Produce a guide to enable statistical analysis of n-of-1 trials.

Methods

The project comprised of two stages: a review of the literature and the development of the key points to consider.

Review

A review was conducted of n-of-1 trials that updated a previous review of the characteristics of these studies (7). The DIAMOND review extended this work to May 2021 by searching published literature and clinical trial registries (ISRCTN (International Standard Randomised Controlled Trial Number) and ClinicalTrials.gov) for abstracts/titles mentioning the phrases "n-of-1". The titles, abstracts and full texts of the identified articles were screened, and those that did not report on a randomised n-of-1 trial were excluded. Data were extracted from the articles and analysed descriptively.

Development of key points to consider, training courses and a statistical analysis guide

In order to develop the key points, firstly a workshop was undertaken where the perspectives of a range of stakeholders were sought, including patient representatives and those who have experience or an interest in undertaking n-of-1 trials. The results from the review were presented, and a discussion facilitated regarding the type of questions that the study design can be used to address, the treatments that can be investigated, and the possible outcomes. Following the workshop, discussions between study collaborators were facilitated to develop a set of key points. The proposed key points were then reviewed by stakeholders at a dissemination event.

The development of the training courses and statistical analysis guide was informed by both the key points themselves and the discussions had during their development.

Results

Review

Fifty-two studies were included in the review, of which (n=8, 15.4%) were in rare diseases. There were no discernible differences in the design characteristics of n-of-1 trials between those undertaken in rare and non-rare diseases. The n-of-1 trials were mainly of drug therapies (n=35, 67.3%). Studies most commonly compared two treatment conditions (n=43, 82.7%) and had a median of six study periods (n=22, 43.3%).

Most of the included studies were placebo controlled (n=34, 65.4%), whilst others used active comparators (n=12, 23.1%), sham devices (n=4, 7.7%) and no intervention periods (n=4, 7.7%).

The majority of studies blinded participants to treatment allocation (n=42, 80.8%).

The majority of studies used patient reported outcome measures (PROMs, n=38, 73.1%). Other studies used: physiological parameters (n=9, 17.3%); behavioural tests (n=3, 6.8%); lab parameters (n=2, 3.8%) and measures of physical activity (n=2, 3.8%).

The statistical analyses were undertaken using a statistical test (e.g. t-test, regression model) (n=35, 67.3%), visual inspection of graphical data (n=2, 3.8%), Bayesian methods (n=8, 15.4%) and time-series analysis (n=1, 1.9%). Some studies did not use statistical analysis methods (n=8, 15.4%). The majority of studies combined the data from multiple n-of-1 trials (n=32, 61.5%).

Development of key points to consider

The workshop was attended by 13 stakeholders. Of these, nine were clinicians/researchers, two were statisticians and two were Patient and Public Involvement (PPI) representatives. Attendees discussed when it is appropriate to undertake n-of-1 trials, the questions that can be addressed and other design considerations.

Following the workshop the key points were developed. Discussions with collaborators resulted in additional points being added regarding the justification for n-of-1 trials and the importance of tailoring the design of the trial to the research question.

Five individuals from the workshop also attended the dissemination event, including a patient representative, a statistician, and three researchers in rare diseases, two of whom had experience of undertaking an n-of-1 trial. Attendees suggested that a point should be added regarding the need for Patient and Public Involvement (PPI) in n-of-1 trials and suggested that improved links should be made between points that have an overlapping scope.

Key points to consider

These discussions led to the development of 21 key points to consider for designing n-of-1 trials, which are listed below in three sections.

Section 1: When is it appropriate to undertake n-of-1 trials?

n-of-1 trials may be used for the following purposes:

1) To inform decisions about the care of an individual patient.

2) For the assessment of very low volume interventions, such as those in rare (and ultra-rare) diseases.

3) For the investigation of expensive health technologies.

4) They can be applied to study wide range of health technologies, provided that

5) the health technology to be assessed has an onset of effect that can feasibly be observed in a study period, and

6) the health technology to be assessed does not have prolonged carryover effects.

Section 2: What types of questions can n-of-1 trials be used to address?

7) n-of-1 trials can address the following questions:

- Does the health technology work at all?

- Does the health technology work better than the existing health technology?

- Which health technology is best for a particular patient?

- Is the treatment effect consistent or does it vary between patients?

Section 3: Design and analysis considerations

The design and analysis section is further divided into eight subsections:

Choice of outcome

8) The question being addressed will inform the choice of primary outcome. Efficacy outcomes are most often indicated, except where the trial is comparing equally efficacious treatment, in which case a patient preference outcome may be used. Practical considerations such as the time point at which an outcome can be observed will also inform outcome choice.

9) A range of outcome measures can be applied. It is recommended to use both patient reported outcome measures (PROMs) and more objective measures of effect where possible, especially in those trials that are being undertaken to assess the efficacy of an expensive or risky treatment.

10) n-of-1 trials can be used not only to assess the effect of a health technology on a primary efficacy outcome but also other outcomes which are important to the patient.

Choice of comparator

11) It is recommended to compare active treatment conditions where possible.

Target of treatment

12) n-of-1 trials can be used to provide evidence regarding the effect of a health technology on the patient's condition itself, specific symptoms of the condition, the side effects of another treatment, or the patients' satisfaction with the health technology.

Number of treatment conditions and periods

13) n-of-1 trials typically compare two treatment conditions. Designing n-of-1 which compare three or more treatment conditions is associated with practical challenges.

14) The number of study periods in an n-of-1 trials is a trade-off between feasibility and precision.

Blinding

15) n-of-1 trials should be blinded where feasible.

Randomisation

16) Blocked randomisation of treatment allocation is typically recommended.

Analysis

17) An interim analysis may be appropriate. The analysis can be using to indicate whether the trial should be stopped early.

18) Washout periods or active (analytical) washout should be employed if there are likely to be carryover effects of the health technologies under investigation.

19) Clinical, rather than statistical, significance should primarily be used to determine the effect of treatment for a particular patient.

20) Combined analysis of a series of n-of-1 trials can be used to estimate the average treatment effect across all the trials, determine whether these effects are consistent for all of the patients in the study, and estimate the average treatment effect in the population generally.

Patient and Public Involvement

21) Relevant and meaningful PPI should be sought throughout the n-of-1 trial including design and planning; interpretation; dissemination and implementation.

Training courses and statistical packages

The training courses were held on 8th September 2022 (for statisticians) and 26th October 2022 (for non-statisticians).

Statistical packages are available here: www.sheffield.ac.uk/scharr/research/centres/ctru/diamond (date accessed 30 June 22).

Discussion

We have developed a set of key points to consider when designing and undertaking n-of-1 trials. They cover when it is appropriate to undertake an n-of-1 trial, the questions that can be addressed, and design considerations. We have also developed additional resources for researchers in order to aid the analysis of such trials.

Strengths and limitations

The strengths are that the project involved a review of previously undertaken n-of-1 trials from 2011 until 2021, representing an up-to-date audit of all randomised n-of-1 trials conducted over this time-period. The review informed a workshop involving 13 stakeholders in the field of n-of-1 trials from a range of backgrounds and disciplines as well as two patient representatives. However, with only 13 stakeholders, the results may not be generalisable compared to if a larger group of individuals were assembled.

Comparison to other studies

The project has built upon previous work on n-of-1 trials methodology. A detailed report by Kravitz et al. (2014) also discusses key considerations for the design and conduct of n-of-1 trials (29). The report states that it is aimed towards a wide audience (including patients, statisticians, and researchers). In the DIAMOND study, we worked with predominantly UK based stakeholders to develop a set of key points to consider for n-of-1 trials targeted at clinicians and trials methodologists interested in undertaking n-of-1 trials in the UK. Additionally, the DIAMOND project involved the development of key materials to support the analysis of these trials in order to reduce the impact of lack of statistical expertise in the use of n-of-1 trials. The project updated a previous review of n-of-1 trials by Gabler et al. (2011), which involved a comparison of the characteristics of n-of-1 trials in rare and non-rare diseases, between which the Gabler review did not distinguish.

Conclusions

Clinical researchers who are planning an n-of-1 trial should consider the key points and utilise the additional materials that have been developed to assist in the design and analysis of the study.

It is recommended that future work should develop targeted protocols for particular types of n-of-1 trials, which could provide additional support with their design and implementation.

1. Introduction

1.1.   Introduction

It has been argued that within-patient, multi-period, crossover trials (n-of-1 trials) are an underused methodology for the evaluation of new health technologies for patients (8). The DIAMOND project (**D**evelopment of general**is**able **m**eth**o**dology for **n**-of-1 trials **d**elivery for very low volume treatments) was undertaken to facilitate the implementation of n-of-1 trials in the evaluation of low volume treatments, such as those for rare diseases.

1.2.   Aims and Objectives

DIAMOND aimed to facilitate the implementation of n-of-1 trials primarily by developing a set of key points to consider to support their design and providing associated training courses and a statistical analysis guide to assist with their implementation.

Specific objectives were to:

a) Review existing literature on n-of-1 trials,

b) Conduct a stakeholder workshop to gain understanding of the questions that can be addressed using n-of-1 trials, the treatments that can be assessed, and possible trial outcomes,

c) Generate key points to consider when designing n-of-1 trials,

d) Provide two training courses (one for statisticians, one for non-statisticians) to disseminate findings,

e) Produce a statistical analysis guide to enable analysis of n-of-1 trials in statistical packages SAS and R.

1.3.   Stages of the Project

The first stage of the project was a review of the characteristics of previously conducted n-of-1 trials. This extended previous work by Gabler et al. (2011) which identified 108 n-of-1 studies to 2010 (7). Next, a stakeholder workshop was conducted to gain the perspectives of a range of stakeholders on the use and design of n-of-1 trials. The findings from the review and workshop were supplemented with discussions between project collaborators in order to ensure the key points to consider were appropriate.

1.4.   Ethical approval

The project was reviewed by the University of Sheffield's School of Health and Related Research's independent Ethics Committee (application number 039799).

1.5.   Scope

n-of-1 trials are designed to evaluate the effects of a health technology in an individual patient. They are referred to as "n-of-1" as the sample size is one patient. Often, a series of n-of-1 trials will be conducted. This report will refer to such as "n-of-1 studies"; they are made up of a number of n-of-1 trials.

It should be noted that, for a study to be considered "n-of-1", it must primarily assess the effect of a health technology at the level of the individual patient. When a series of n-of-1 trials are conducted, a secondary objective might be to estimate population level effects using meta-analysis. Such studies should be distinguished from group level crossover trials, where a large number of participants are randomised to a sequence of treatments and whose primary aim is estimation of population level effects.

A simple way to ascertain whether a within-patient crossover study is an n-of-1 study is to consider whether the aim of the evaluation is to inform a treatment decision for that particular patient. If it is, then the study can be considered "n-of-1".

The project and its outputs are focussed on n-of-1 trials which randomise the treatment allocation to each period as these are widely favoured due to increased methodological rigour. However, many of the key points could be applied to the design of n-of-1 trials which do not randomise treatment allocation in this way.

1.6. A note on language

Throughout this report, the terms treatment and health technology are used to refer to the subject of the evaluation in an n-of-1 trial, such as a drug or medical device. The term 'health technology' was chosen to emphasise that not all n-of-1 trials evaluate 'treatments'. However, it is acknowledged that the term 'treatment' is sometimes preferable, for example when referring to 'treatment effects'. Thus both terms are used in this report according to context.

It is important to note that the number of health technologies evaluated in an n-of-1 trial refers not only the active health technologies, but also their comparators. For example, an n-of-1 trial evaluating the effects of a drug compared to a placebo would be described as comparing two health technologies. In the same way, a 'no intervention' comparator would count towards the number of health technologies compared.

1.7. Document Roadmap

In the following chapters we provide a background to n-of-1 trials, followed by the methods and results of the DIAMOND project. Finally, we provide a set of key points to consider when designing and implementing n-of-1 trials in chapter 6, followed by a discussion of the results and future implications.

The following report is set out as follows:

- **Chapter 2: Background to n-of-1 trials.** The need for n-of-1 trials and the view that they are an underused design is outlined. Key literature and terms are summarised and potential reasons behind this underuse are highlighted.
- **Chapter 3: Statistical analysis guide.** The statistical analysis both of individual n-of-1 trials and studies conducted as a series of n-of-1 trials is described.
- **Chapter 4: Review of previously conducted n-of-1 trials.** A review of the design characteristics of n-of-1 trials since 2011 is reported upon.
- **Chapter 5: Development of a set of key points to consider.** The steps taken to develop key points to consider for the design, implementation and analysis of n-of-1 trials are summarised. These were a workshop with key stakeholders (n-of-1 methodologists, patient representatives, clinicians), a discussion between the study team and collaborators, and a stakeholder dissemination and feedback event.

- **Chapter 6: Key points to consider.** A set of key points to consider for the design of n-of-1 trials are presented. Case studies are used to exemplify the points.
- **Chapter 7: Discussion**. The main findings and implications of the DIAMOND project are discussed.

2. Background

2.1. The need for n-of-1 trials

When it comes to establishing that treatments have general efficacy, randomised controlled trials (RCTs) are widely considered the gold standard approach (9). Most such trials are parallel group trials. These use randomisation to determine which intervention a patient receives, ensuring there are no systematic differences between the groups of patients receiving each intervention and enabling the effects of unseen random differences to be estimated and appropriately reflected in standard errors and confidence intervals. This is important so that the extent to which any differences in outcomes between the groups can be attributed to the interventions they received may be assessed and uncertainties regarding such assessment can be expressed.

Although these parallel-group RCTs cannot recruit a sample of patients that is perfectly comparable to target populations, a combination of judicious scales of measurement and background knowledge regarding relevant factors can be used to establish that they are generally useful. This is valuable for making inferences about the general utility of a treatment. However, it is not unreasonable to suppose that one could do even better by making more nuanced judgements as to when and for whom treatments should be used. A typical approach to do this is to examine particular subgroups defined by characteristics that might plausibly affect the response to treatment. However, such an approach inevitably lacks precision since patients in the subgroups are fewer in number and sometimes much fewer than those in the trial as a whole. In any case, the approach is indirect. Where practical circumstances permit it a direct 'trial of therapy' is a possibility.

'Trials of therapy' are informal evaluations of the effects of treatment on an individual patient and are a common tool used to inform decision making in routine clinical practice (1). They typically involve a patient 'trying out' a treatment for a period of time. If the treatment is deemed effective, the patient may continue to receive this treatment. If there is no perceived benefit of the treatment, the clinician may switch the patient to an alternative. However, without controlling for confounding variables, these evaluations will be prone to bias. As a result, these evaluations may be of limited use in determining the optimal treatment for a particular patient.

n-of-1 trials are a type of crossover trial designed to optimise the evaluation of health technologies in individual patients. In these trials, a single patient receives one treatment for a period of time and then switches to receive another (10). Whilst receiving each treatment, outcomes will be assessed using appropriate measures in order that the effects of the treatment(s) under investigation can be established. The 'switching' is repeated, such that multiple measurements are obtained for each treatment. Usually, the order in which the patient receives the treatments is randomised. These n-of-1 trials are RCTs conducted in individual patients. Such n-of-1 trials combine methodological rigour with the investigation of treatment effects in an individual patient to enable the determination of the optimal treatment for a particular patient.

n-of-1 trials can be employed to study health technologies for conditions and treatments for which parallel-group RCT evidence is not available. Parallel group trials require a large sample of patients to study. This is not possible in all conditions. Rare diseases are those which affect fewer than one in 2,000 people according to the European Union definition (11). If there is not a large patient population from which to sample, then it is difficult to conduct a parallel-group trial. n-of-1 trials do not require this and therefore could be implemented to research such conditions. Further to this, rare diseases

are typically highly heterogeneous, making the investigation of individual responses to treatment more valuable. Thus, n-of-1 trials might be an appropriate and practical way to study treatments for rare diseases (4,12). The burden of rare diseases is high; there are an estimated 5,000 – 8,000 rare diseases affecting an estimated 6.2% of the general population (13,14). The implementation of n-of-1 trials in rare diseases could be one way of addressing this.

Table 1: Key terms for n-of-1 trials

Term	Definition
Period	The duration of time in which the patient receives one of the health technologies being evaluated.
Crossover	The switch of treatment from one period to another.
Cycle	A set of periods in which the participant receives all of the health technologies being evaluated.
Pair	A cycle comprised of two periods.
Block	A set of episodes of treatment for which random allocation has been constrained. For example if a two-treatment design comparing (say) A & B is blocked in pairs of periods then in each pair of treatments A will occur and so will B but the order will be random. If a block of four periods were chosen two As and two Bs will appear in every four periods but the order will be random.
Washout	A period of time following a study period during which the participant receives no treatment. This is to ensure that any effects of the previous treatment do not carry over into the subsequent study period.
Active washout	The patient switches from one study period to the next, without a washout period. Outcomes are not assessed (or data from such assessments are ignored) for a period of time in order that the effects from the previous treatments have washed out before assessment of outcomes for the subsequent study period begins.
Randomisation	The treatment allocation to study periods is determined by chance.
Health technology	A broad term encompassing any treatment or intervention which might be evaluated using an n-of-1 trial e.g. a drug, behavioural intervention or medical device.
Placebo	A substance that has no therapeutic effect and is used as a control when testing a new drug.
Repetition	n-of-1 trials are comprised of multiple cycles of study periods. This is to minimise the effect of random error (where the outcomes are affected by extraneous variables).
Blinding	Where one or more parties are kept unaware of which treatment (e.g. active drug or a placebo) the participant is receiving at any given time. Clinicians, patients or outcome assessors can be blinded.

2.2. n-of-1 trials in the literature

n-of-1 trials were first introduced into the medical literature in 1986 by Gordon Guyatt and his colleagues at McMaster University, Canada, who described them as double-blind, randomised trials in which a single patient is allocated a series of treatments, typically a drug and placebo allocated in pairs of periods, the order being random (10). Variations on this design have since been described and are apparent in the literature.

A systematic review by Gabler *et al.* (2011) captured the characteristics of 108 randomised n-of-1 trials published since Guyatt et al.'s seminal paper (7). Although n-of-1 trials primarily assessed drug therapies in this period, they were also used to assess behavioural, surgical, and medical device interventions in myriad conditions. They compared a median of two health technologies (e.g. drug and placebo, range 2-6). Data were analysed using a variety of approaches, including comparison with specified statistical significance levels and visual examination of graphical data. There was high variability in the quality of reporting of n-of-1 trials, however, with some providing sufficient information to enable meta-analysis or to determine the outcome of the trial, for example, whether the patient's treatment changed as a result of the trial findings. The n-of-1 trials captured in this review were varied in both their design and reporting quality and similar findings have since been reported elsewhere (15).

In response to evidence of inadequate reporting of n-of-1 trials, the CONSORT extension for reporting n-of-1 trials (CENT) guidelines were developed (16). They include a checklist of items, adapted from the CONSORT 2010 checklist (17), which should be reported. These include items such as identifying the study as an n-of-1 trial in the title, outlining the rationale for using the n-of-1 design, and describing the method used to generate the sequence for treatment allocation. These guidelines were developed with the aim of improving the completeness of reporting of n-of-1 trials and were expected to be of use to authors of n-of-1 trials and to the reviewers who assess them (16).

To date there has been one published review of n-of-1 studies in rare diseases (18). Müller et al. (2021) reviewed n-of-1 studies in rare genetic neurodevelopmental disorders such as Fragile X Syndrome and Down's Syndrome. Twelve studies were identified, published between 1978 and 2017. Although they included randomisation in their definition of an n-of-1 trial, both randomised and non-randomised designs were eligible for inclusion. The studies had a mean of five participants and had wide variation in design elements such as number of periods and total study length. The included studies were assessed against the CENT criteria and vide variation in reporting quality was apparent. Nine of the studies had a predefined primary outcome measure, however it was only explicitly described as such in three studies. There was variation in how the results of the n-of-1 studies were interpreted. In nine of the studies, results were interpreted as generalizable to the population whereas the authors of two studies considered the trial findings as evidence applicable to the trial participant(s) only. One study did not report on interpretation of trial findings. Strengths of the included studies were identified as the intent to measure personalised and clinically relevant outcomes and the use of both subjective and objective measures of treatment effect. The authors identified difficulty with the statistical analysis of n-of-1 trials as a limitation of the studies. This review highlighted that, whilst being an area well suited to the study design, the use of n-of-1 trials has been limited in rare genetic neurodevelopmental disorders.

A user's guide to the design and implementation of n-of-1 trials was commissioned by the agency for Healthcare Research and Quality in the US (19). This document aimed to improved understanding of n-of-1 trials and implementation in a very broad audience. It provides guidance and information on a range of issues and considerations for n-of- trials.

2.3.   An underused design

n-of-1 trials remain an underused design for the evaluation of health technologies in individual patients (8). It is worth considering the possible reasons behind this underuse, in order that they may be addressed to facilitate the wider implementation of n-of-1 trials.

Richard Kravitz examined this question by consulting some of the key figures in the n-of-1 movement (6). He highlighted a range of potentially contributing factors, including conceptual confusion about whether n-of-1 trials ought to be considered a research tool, or an approach to optimising routine clinical care. n-of-1 services have been some of the most successful approaches to n-of-1 implementation and these tend to align themselves with the view of n-of-1 trials as a clinical service involving research methods (1). These research methods, including randomisation and blinding, may mean the trials must comply with research governance. Many of the challenges associated with implementing n-of-1 trials in routine care are structural and regulatory (20). The challenges associated with n-of-1 trials as a research tool might be more amenable to influence through education and guidance, for example, clinicians not believing in the clinical value of n-of-1 trials; lack of understanding of the technical and statistical elements of the design (6). If these are factors contributing to the underuse of n-of-1 trials, then it might be possible to increase their implementation but improving understanding of their design and methods.

2.4. The present study

The DIAMOND study addresses the problem of n-of-1 trials being underutilised, particularly with respect to their potential in rare disease research. It provides guidance in the form of a set of key points to consider for the design and use of n-of-1 trials with associated case studies to provide examples in the literature and freely available training and materials to support the statistical analysis of n-of-1 trials. Importantly, the key points to consider were developed alongside key stakeholders to help ensure they are practicable and acceptable.

3. Statistical analysis guide

3.1. Introduction

This chapter may be used as a guide to the statistical analysis of n-of-1 studies. It begins by describing a study undertaken in a single patient before extending the analysis to where a series of n-of-1 trials have been carried out.

3.2. Randomisation

n-of-1 studies comparing two treatments are often designed such that, within a given pair of periods, each treatment is used once (21–23). These pairs of periods are referred to as *cycles* (24).

With this design, a study comparing two treatments over six periods would comprise of three cycles. Table 2 gives all possible sequences for a study designed in this way. If the study involved only one patient, then that patient would be allocated at random to one of the eight possible sequences.

Table 2. Set of sequences for a design using six periods arranged in three cycles.

	Periods					
	1	2	3	4	5	6
Sequence	Cycle 1		Cycle2		Cycle 3	
1	A	B	A	B	A	B
2	B	A	A	B	A	B
3	A	B	B	A	A	B
4	B	A	B	A	A	B
5	A	B	A	B	B	A
6	B	A	A	B	B	A
7	A	B	B	A	B	A
8	B	A	A	A	B	A

If there are k possible cycles in which patients can be treated, there will be 2^k possible sequences. The advantage of this design is that after two periods a patient will have experienced both treatments which is informative. It also prevents a sequence in the form of AAABBB where a patient may not experience the second treatment ('B') until the fourth period (in a 6-period design).

If a series of n-of-1 trials are being carried out, then blocking could be introduced so that, for example, with a block size of 8, for every 8 patients, each would be allocated to one of the eight sequences at random, all 8 sequences being used. The disadvantage of such blocking would be that the number of patients recruited would have to be a multiple of 8.

3.3. Analysis of an n-of-1 trial

For an individual n-of-1 trial, the analysis is often descriptive. A useful plot of the data is given in Figure 1. The data are from patient 1 in Table 4 later in the chapter. The figure gives the plotted paired treatment data for each cycle, showing that the outcome for treatment B is consistently higher than that for treatment A for this patient.

To assist in the interpretation of such data, it could be informative to define clinically meaningful effects for an outcome above which it could be concluded there is a treatment effect.

Figure 1. Paired treatment effects by cycle.
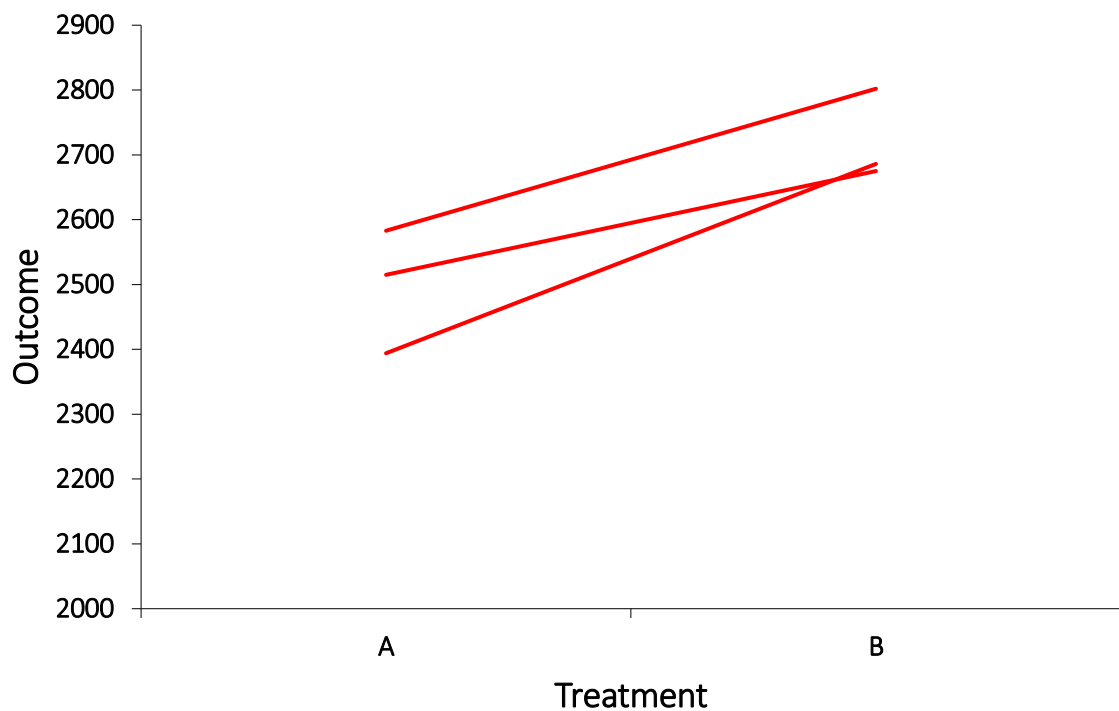
A further informative plot is Figure 2. Here, the result for each cycle is represented by a red circle plotting the value under B (y-axis) against that under A (x-axis). The diagonal line represents equality between the two treatments. The average value over all cycles is given by a black square.

The figure is complementary to Figure 1. The red circles are generally above and to the left of the line of equality suggesting that B has an effect bigger than A.

For both figures the analysis could be repeated across several clinical outcomes to help to quantify the effect of treatment.

Figure 2. Treatment responses by cycle.
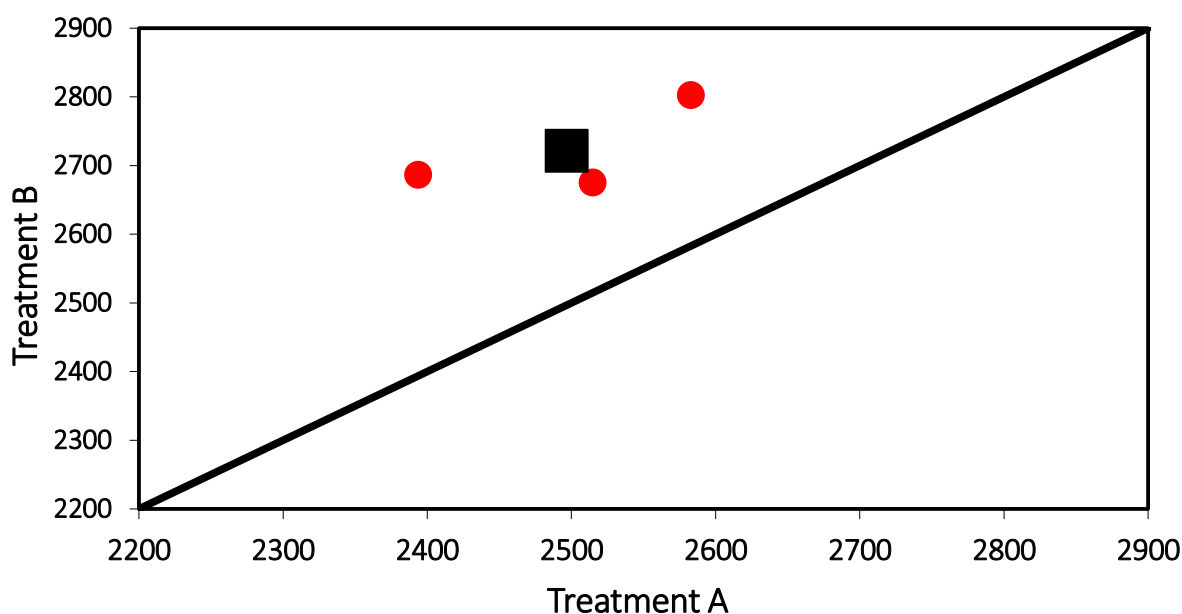
3.3.1. Interim summary

Analysis of a single n-of-1 trial is often descriptive and might be conducted across a number of clinical outcomes to help determine whether there is an effect of treatment. As far as possible, the analysis of the study should be prespecified along with definitions of clinically meaningful effect sizes.

3.4. Analysis of several trials

A number of different questions can be addressed when a series of n-of-1 trials have been undertaken (25).

Table 3. Research questions from n-of-1 trials

1.	What was the effect for individual patients in the trials?	This question would be initially addressed by undertaking analyses for each individual n-of-1 study but can also be revisited when analysing data across patients.
2.	Was there an effect of treatment in the trials?	As well as if there is an effect, is it a clinically meaningful effect?
3.	Was the treatment effect identical for all patients in the trials?	If the effect is not identical is there a subgroup of patients for whom there is an effect?
4.	What was the average effect of treatment in the trials that were run?	Given point 3. this average could be across all patients or across a subgroup of patients
5.	What will be the effect of treatment when used more generally (in future)?	This will inform treatment decisions for patient population in whom the n-of-1 trials have been undertaken.

The following analyses will be organised according to these questions, using simulated data presented by Araujo et al (26) for illustrative purposes. In the imagined data set 12 patients have been

randomised in three cycles to treatment A followed by B or B followed by A. The table gives the periods in which the patients received A or B. For example, Patient 1 received treatment A in periods 1, 3 and 6 and treatment B in periods 2, 4 and 5. The first ten patients have completed all three cycles of treatment. However, patient 11 has only completed two cycles of treatment and patient 12 has only completed 1. This has been done to illustrate a complication in analysis that may arise in practice. We thus have data from $(10\times3)+2+1=33$ cycles and therefore from $2\times33=66$ periods.

Table 4. A simulated set of n-of-1 trials of FEV$_1$ (mL) outcome data.

Patient	Treatment A	B	A	B	A	B
1	1	2	3	4	6	5
	2394	2686	2515	2675	2583	2802
2	2	1	3	4	6	5
	2746	2726	2592	2867	2743	2742
3	1	2	3	4	6	5
	2668	2560	2542	2584	2491	2737
4	1	2	3	4	6	5
	2397	2696	2411	2895	2499	2760
5	2	1	3	4	5	6
	3179	3221	2952	3096	2600	3192
6	1	2	4	3	5	6
	2643	2496	2759	2847	2651	2860
7	1	2	3	4	5	6
	2678	2843	2492	2763	2801	2890
8	2	1	3	4	5	6
	2887	2862	2875	3083	2689	2967
9	2	1	3	4	6	5
	2490	2841	2648	3044	2688	2914
10	2	1	3	4	6	5
	2268	2576	2413	2493	2344	2699
11	2	1	4	3	6	5
	2617	2923	2629	2832		
12	1	2	4	3	5	6
	2627	2759				

The first step is to consider the individual trials. For example, Figure 1 could be repeated through a trellis plot. Figure 3 presents the data from the first nine patients from Table 4. After analysing the data descriptively, more formal analysis could be undertaken. This assessment will help to answer the first three questions from Table 3.

Figure 3. Paired treatment responses by patient.
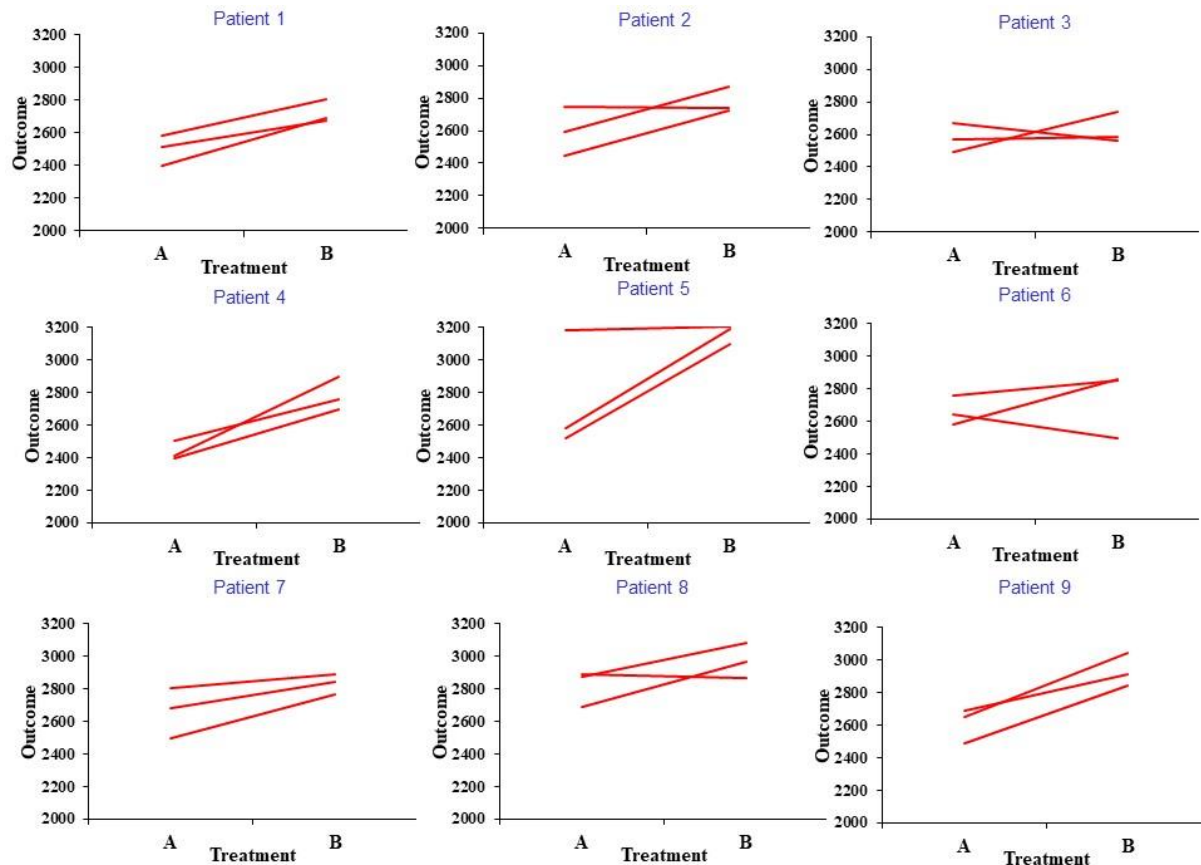
3.4.1. Demonstrating that there can be a difference between treatments

The relevant null hypothesis here is that there is no difference between treatments for any of the patients. If that is the case, under the null hypothesis, it does not matter which patient is studied, the result may be expected to be the same. Thus, a matched pair analysis on all 33 cycles can be carried out.

The data have been reduced to differences by cycle and patient and are presented in Table 5. The differences can be analyses by a one-sample t-test for which the statistics in Table 6 are produced.

The calculations for Table 5 and Table 6 are given in the Excel file detailed in Table 13 in the sheets 'Data' and 'By Cycle'.

Table 5. Differences (Treatment B - Treatment A) per cycle arranged by patient

	Cycle		
Patient	1	2	3
1	292	160	219
2	-20	275	-1
3	-108	42	246
4	299	484	261
5	42	144	592
6	-147	88	209
7	165	271	89
8	-25	208	278
9	351	396	226
10	308	80	355
11	306	203	.
12	132	.	.

Table 6. Summary statistics for a one-sample t-test based on differences per cycle.

Statistic	Value	Explanation
n	33	Number of cycles
Mean	194.55 mL	Mean of the 33 cycles
95% CI	137.2 to 251.9	Plausible range for the mean response
P-value	<0.001	Probability under H_0 a t-statistic with 32 DF will be ≥ 6.91 or ≤ -6.91

This particular analysis can be criticised. This is because it is treating each data point as an estimate of effect such that there are more data points in the analysis than there are people and, as a consequence, more degrees of freedom than people.

The calculations in Table 6 can be criticised. Whilst it is reasonable to assume by hypothesis that the treatment effect is constant for all patients when we are testing that this effect is zero for them all, as soon as we allow that the effect is *not* zero, it becomes plausible that it might vary from patient to patient. If we regard the patients as being fixed, that is to say that we are only making a statement about these patients, then we could claim that this source of variation would not contribute to the treatment estimate changing were we to repeat the experiment. However it will contribute to the overall estimate of variation that we have used.

This source of variation can be eliminated by constructing variance estimates patient by patient. The calculations are given in Table 7.

Table 7. Intermediate calculation for estimating the common within-patient variance (note that the units of variances and sums of squares are mL^2 of FEV$_1$).

Patient	DF	Variance	Sum of Squares
1	2	4372.3	8744.7
2	2	27260.3	54520.7
3	2	31572	63144
4	2	14233	28466
5	2	85601.3	171202.7
6	2	32767	65534
7	2	8356	16712
8	2	25166.3	50332.7
9	2	7758.3	15516.7
10	2	21636,3	43272.7
11	1	5304.5	5304.5
12	0	0	0
Total	21		522750.5

Here the column labelled DF gives the degrees of freedom patient by patient and is equal to the number of cycles minus 1. The column labelled Variance gives the *local* estimate of the variance of the differences (B-A) patient by patient. For patient 12 the value is zero since the patient was only studied in one cycle and hence there is only one difference. The column labelled Sum of Squares is obtained by multiplying the variance by the degrees of freedom. The overall sum of squares is 522,750.5 and if this is divided by the total DF, 21, we obtain 24,893, which is thus out estimate of the variance on the assumption that variability does not vary from patient to patient. The standard error is 157.8.

The consequent calculations are summarised in Table 8. These calculations are given in the Excel file detailed in Table 13 in the sheets 'Data' and 'By Patient'.

Table 8. Summary statistics to perform a one-sample t-test based on differences per cycle with the patient by treatment interaction removed from the variance estimate.

Statistic	Value	Explanation
n	33	Number of cycles
Mean	194.55 mL	Mean of the 33 cycles
95% CI	137.4 to 251.7	Plausible range for the mean response
P-value	<0.001	Probability under H$_0$ a t-statistic with 32 DF will be ≥ 7.08 or ≤ -7.08

A criticism of this analysis is that, if we do not regard the patients as fixed then we have not reflected the variation from patient to patient enough, since our estimate is based on using cycles as the unit of inference rather than patients.

3.4.2. Putting more general bounds on the treatment effect

One way of proceeding is to reduce the differences to a mean per patient and then perform an analysis using these 12 means differences as our raw input. The data are presented in TABLE X. We shall ignore

the column labelled standard error for the moment. (We shall use this later.) Instead, we just base our analysis on the 12 *per patient estimates*.

Table 9. Mean per patient.

223.7
84.7
60.0
348.0
259.3
50.0
175.0
153.7
324.3
247.7
254.5
132.0

If we carry out a one-sample t analysis on these values, we can summarise the results as in Table 10. Summary statistics for a one-sample t-test based on differences per patient.

The calculations in Table 10 are given in the Excel filed detailed in Table 13 in the sheets 'Data' and 'By Patient'.

Table 10. Summary statistics for a one-sample t-test based on differences per patient.

Statistic	Value	Explanation
n	12	Number of patients
Mean	192.74 mL	Mean of the 12 patient means
95% CI	129.5, 255.9	Plausible range for the mean response
P-value	<0.001	Probability under H_0 a t-statistic with 11 DF will be ≥ 6.71 or $\leq$-6.71

The result from this analysis is similar to that reached before. However, the analysis is conceptually different to that previously provided as it is relevant to a different question: what can we say about the mean effect in general, not just for patients studied?

3.4.3. Meta-analytic approaches

A set of n-of-1 trials such as we have been considering is analogous to a collection of results from independent clinical trials, such as might be summarised in a meta-analysis. Existing tools for meta-analysis can be adapted to perform the analysis of a set of n-of-1 trials.

There is one important change in data-preparation that is, however, necessary. Standard meta-analytic approaches assume that the standard errors used to calculate the weights are themselves calculated without error. This is not true as estimated standard errors are random variables not known parameters. However, if the associated degrees of freedom are reasonably large, this assumption does not matter. For n-of-1 trials, however, the degrees of freedom are typically small. In our example, there are no more than two per patient. Naively estimating the variances independently is unwise (27). It is better to use a pooled variance to do so.

Thus, we impose an assumption that the within-patient *variation* between estimates per cycle is constant across patients. We then proceed to estimate the variance. The approach is illustrated in Table 11.

For each patient the degrees of freedom are calculated as the number of cycles in which they were treated minus one. The values are shown in column two. The sample variance of the estimated treatment effect for each patient is calculated and given in column three. The product of the values in columns two and three gives the sums of squares (corrected by the mean), which is shown in column four. (If the available statistical software package has a standard function available for the *corrected sum of squares*, it may be easier simply to calculate column four directly.) The sum of the values in column four is 522750.5 mL^2. Dividing the total sum of squares by the total degrees of freedom, 21, yields an estimated variance of 24892.9 mL^2 and the square root of this is 157.77 mL

Table 11. Intermediate calculation to estimate the common within-patient variance. (Note that the units of variances and sums of squares are mL^2 of FEV$_1$).

Patient	DF	Variance	Sum of Squares	Standard Error
1	2	795	8744.7	91.09
2	2	4956.4	54520.7	91.09
3	2	5740.4	63144	91.09
4	2	2587.8	28466	91.09
5	2	15563.9	171202.7	91.09
6	2	5957.6	65534	91.09
7	2	1519.3	16712	91.09
8	2	4575.7	50332.7	91.09
9	2	1410.6	15516.7	91.09
10	2	3933.9	43272.7	91.09
11	1	482.2	5304.5	111.56
12	0	0	0	157.77
Total	21		522750.5	

Note that since patient 12 was only treated in one cycle, it is impossible to estimate a variance for them. However, using the data from other patients we assume that the estimated standard deviation for them is the same as for all patients and is thus 157.8 mL. Since the estimate for this patient is only based on one cycle, the standard error for them is the same as the standard deviation since $157.77 \; mL/\sqrt{1} = 157.77 \; mL$.

In general if a patient was treated in k cycles, we have $SE = s/\sqrt{k}$ where *s* is the estimated pooled standard deviation ($157.77 \; mL$ for this example). For patient 11 we have *k = 2* and for all other patients *k = 3*. Substituting these values of *k* yields the standard errors given in Table 12.

Table 12. Summary statistics per patient for a meta-analysis.

Per patient estimate	Standard error
223.7	91.09
84.7	91.09
60.0	91.09
348.0	91.09
259.3	91.09
50.0	91.09
175.0	91.09
153.7	91.09
324.3	91.09
247.7	91.09
254.5	111.56
132.0	157.77

We can now apply standard meta-analytic approaches to the data in Table 12. Figure 4 illustrates the analysis using Guido Schwarzer's **meta** package (28).

Figure 4. Results of a meta-analysis.
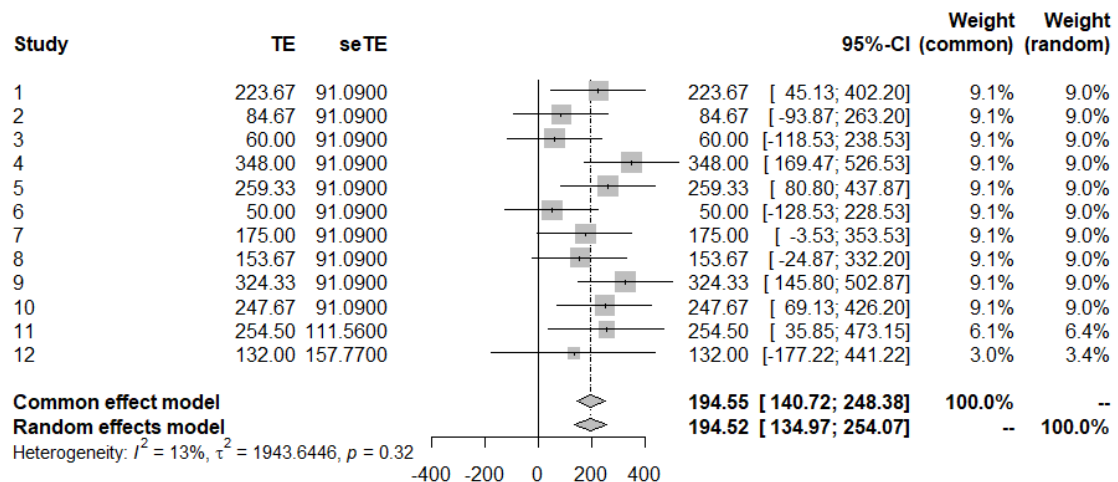
This provides both a fixed and a random effects analysis. For this example, the results are very similar. Furthermore, the point estimate of 194.55 is identical to that reached for the matched pairs analysis of the 33 cycles. This is because the standard errors patient by patient have been calculated using the same variance, the difference between them merely reflects the numbers of cycles for which information was obtained.

The standard error is different, however. This is based on 21 degrees of freedom rather than 32. The extent to which results vary from patient to patient have been removed from the estimate of the variance. The difference is 11 degrees of freedom and these are the degrees of freedom that correspond to the treatment-by-patient interaction.

The random effects meta-analysis estimate has a slightly wider confidence interval. This is because it provides an estimate of the treatment effect that would apply were it the case that the patients that have been studied were no longer fixed but could be regarded as a random sample from a wider but 'similar' population. Thus the terms that the interaction measures are no longer regarded as being fixed but values that might be different from one occasion to another. Thus this uncertainty is incorporated in the confidence intervals.

In favour of the random effects analysis is the fact that it addresses a more important question. Against it is the fact that strong assumptions (the similarity of patients studied with those in the target population) have to be made.

3.4.4. How meta-analysis can inform decisions for an individual patient

Even if there were a treatment effect with a meat difference of µ between treatments, not every patient would observe this effect – half of the patients would see an effect greater than µ and half smaller than µ. In fact, if the alternative hypothesis is true, we would expect to see a range of responses with µ being the average effect.

Suppose patients 1, 2, 3, and 4 are a subpopulation of patients – presented in Figure 3– whom we wish to analyse. Looking at the individual data there does seem to be an effect for patients 1, 2 and 4 but the evidence for patient 3 is equivocal.

If there is no reason to believe patient 3 to be different to the other patients then the results from the meta-analysis could be used to inform the treatment decisions for that patient. Thus, there would be benefit for individual patients from the meta-analysis.

3.4.5. Estimates of effect for individual patients

It is possible that superior estimates of effect for individual patients can be obtained by using the results from others. This is because a series of n-of-1 trials will provide two sorts of information for a given individual, namely *personal* and *global*, the former using only a given patient's data and the latter all of the data. Each of these is an unbiased estimate of the effect for a patient and they may be combined to produce a so-called *shrunk* estimate as follows

$$shrunk = w \times personal + (1-w) global$$

Where *w* is a weight between 0 and 1. The greater the value of *w*, the greater attention we pay to the information from the given patient. The estimate is referred to as shrunk because the result will lie between personal and global and so may be regarded as having shrunk towards the latter compared to the former. An alternative term is *best linear unbiased predictor* (BLUP).

Just as we combine information from a meta-analysis by weighting the trial proportionately to the inverse of the variances of their estimates, so we weight these two sorts of information inversely according to their variances.

Fortunately, this sort of question is addressed in various meta-analytic packages. For example, the metafor package (29) within R has a blup( ) function that will do this.

3.5. Use of programs

Programs have been provided in SAS, Genstat and R which undertake the fixed and random effects meta-analysis methods described in this chapter. There are also programs for undertaking sample size calculations across a series of n-of-1 trials which have not been described. Descriptions of the programs are given in Table 13. Statistical programs for the analysis of n-of-1 trials..

Table 13. Statistical programs for the analysis of n-of-1 trials.

Package	Program	Description
Excel	Excel Tutorial Calculations	The document undertakes the calculations illustrated in Table 5, Table 6, Table 8 and Table 10.
SAS	Diamond example analysis mixed balance	This program analyses the data using the methods described in the chapter. It uses a data set for which there is complete data for all the patients (6 treatment periods and 3 cycles).
	Diamond Example analysis mixed unbalanced	This program analyses the data using the methods described in the chapter. The data set does not need to be complete data and the data set used as an example is the same as that in the chapter.
	Diamond sample size	This program produces sample size calculations for estimating the overall mean from a set of n of 1 trials. Fixed and random effects methods are undertaken.
R	Diamond N of 1 linear model analysis	This program does the same calculations as "Diamond example analysis mixed balance" in SAS.
	Diamond N of 1 linear model analysis unbalanced	This program does the same calculations as "Diamond example analysis mixed unbalanced" in SAS.
	Diamond N of 1 sample size	This program does the same calculations as "Diamond sample size" in SAS.
GenStat	Diamond n of 1 simulation analysis	This program does the same calculations as "Diamond example analysis mixed balance" and "Diamond example analysis mixed unbalanced" in SAS. It also does additional analyses including trellis plots.
	Diamond n of 1 power	This program does the same calculations as "Diamond sample size" in SAS.

3.6. Summary

This chapter has described the methods for the analysis of n-of-1 studies from the simple case of a single n-of-1 trial to combining a series of n-of-1 trials in a meta-analysis.

4. Review

4.1. Introduction

A review of the characteristics of n-of-1 trials was conducted, updating previous work by Gabler et al. (2011) (7).The review was later used to guide the development of key points to consider for the design and implementation of n-of-1 trials in the stakeholder workshop.

4.1.1. Aims

- To provide an overview of the characteristics of published n-of-1 trials between 2011 and 2021.

- To understand whether the characteristics of n-of-1 trials in rare disease areas differ from those in non-rare disease areas.

4.2. Methods

4.2.1. Trial identification

Searches were conducted on 5th May 2021 using PubMed, EMBASE, Web of Science, the NIHR Journals library and clinical trials registries (ISRCTN and ClinicalTrials.gov). The term "n-of-1" was searched in the "Title" field and a filter was applied to retrieve results published since 2011. It was agreed that using the search term "n-of-1" only would capture the relevant results.

The titles and abstracts of the resulting articles were screened, and those that did not report on an n-of-1 trial were excluded. Full text screening for eligibility was conducted by one reviewer [OH] and a random 10% of these were independently checked by a second reviewer [RC]. If there were any questions on a study from OH, then they were discussed with SAJ.

The following eligibility criteria were used:

a) The article had to report on an interventional study in a human population

b) The article had to present the protocol for or the results of a study.

c) The order of treatment episodes had to be randomised, where the unit of randomisation was the individual treatment episodes

d) Where multiple n-of-1 trials were conducted within a study, the primary analysis had to be at the level of the individual

e) Abstracts were included where no full text was available, given they included sufficient information for data extraction.

f) Papers in languages other than English were excluded unless an English abstract meeting criterion (e) was available.

g) Reviews of n-of-1 studies were excluded.

4.2.2. Data Extraction

Data were extracted by one reviewer [OH] into Microsoft Access (2016) and a random 10% of these were independently checked by a second reviewer [RC]. Variables for extraction were adapted from Gabler et al. (2011) and equivalent categories were used where appropriate. Where additional

variables were extracted, categories were developed by OH. If there were any questions, these were discussed with RC and SAJ until consensus was reached.

The variables for data extraction were:

General characteristics: Publication year, funding, region, type of publication (e.g. protocol).

Sample characteristics: Target sample size (for protocols only), number of patients randomised, number of patients completing, age of participants.

Design and intervention characteristics: Intervention type, health/disease area, rare disease, number of treatments, number of periods, period length, number of crossovers, number of cycles, washout, length of washout, randomisation in pairs, type of comparator, blinding, total duration.

Outcome characteristics: Method of primary outcome measurement, frequency of outcome measurement, definition of response to treatment, treatment change subsequent to trial completion.

Analysis characteristics: Type of analysis, meta-analysis, reporting of numerical results, p values, and significance level.

A full list of variables and categories is available in appendix 1.

4.2.2. Data Analysis

Descriptive statistics were produced for all variables. Discrete variables were expressed as number and percentages, and continuous variables were expressed as medians, modes and ranges. Missing data were coded as 'not reported' for discrete variables and excluded from the analysis of continuous variables.

4.3. Results

Searches identified 1416 records and, after the removal of duplicates, 869 unique records were screened. Of these, 243 were sought for retrieval and 52 records were found to be eligible for inclusion. No additional records were identified by hand searching the references of nine reviews of n-of-1 trials.

Figure 5. Selection of reports (30)

Identification of studies via other methods

Records identified from:
Citation searching of identified reviews (n = 0)

Reports sought for retrieval (n = 0)

Reports assessed for eligibility (n = 0)

Reports not retrieved (n = 0)

Reports excluded (n=0)

Identification of studies via databases and registers

Records identified from:
Databases (n =1416)
Registers (n = 6)

Records removed *before screening*:
Duplicate records removed (n = 547)

Records screened (n = 869)

Records excluded (n = 626)

Reports sought for retrieval (n = 243)

Reports not retrieved (n = 4)

Reports assessed for eligibility (n = 239)

Reports excluded:
Not an interventional study (n = 50)
Non-human population (n = 2)
Treatment episode order not unit of randomisation (n = 83)
Insufficient information (n = 11)
Text not in English (n = 1)
Review paper (n = 9)
Outcome of interest not ITE (n =30)

Studies included in review (n = 52)
Reports of included studies (n = 52)

Identification

Screening

Included

4.3.1. General characteristics

Eight studies (15.4%) were in rare diseases and 44 (84.6%) were in non-rare diseases. 12 (23.1%) were protocols for n-of-1 studies and the remainder (n=40, 76.9%) reported on the results of completed studies.

The majority of studies reported receiving funding (n=32, 61.5%). Eight studies (15.4%) did not receive any funding, all of which were in non-rare disease areas. They spanned a range of health and disease areas, with the most common being neuropsychiatric conditions (n=16, 30.8%). Four studies (7.7%) were not conducted in a specific disease area e.g., blood transfusion dependent patients (31,32)

The studies were conducted in a range of countries. These included, but were not limited to, the UK (n=5. 9.6%), the Netherlands (n=6, 11.5%), China (n=7, 13.5%), Australia (n=7, 13.5%), USA (n=9, 17.3%) and Canada (n=6, 11.5%). Most of the studies in rare disease areas were conducted in the Netherlands (n=5, 62.5%).

Table 14. General characteristics of n-of-1 studies.

Variable		Rare n	Rare %	Non-rare n	Non-rare %	Total n	Total %
Disease type							
Protocol	Yes	2	25.0	10	22.7	12	23.1
	No	6	75.0	34	77.3	40	76.9
Funding	Yes	6	75.0	26	59.1	32	61.5
	No	0	0.0	8	18.2	8	15.4
	Not Given	2	25.0	10	22.7	12	23.1
Region	UK	0	0.0	5	11.4	5	9.6
	Netherlands	5	62.5	1	2.3	6	11.5
	France	1	12.5	2	4.5	3	5.8
	China	1	12.5	6	13.6	7	13.5
	Australia	0	0.0	7	15.9	7	13.5
	USA	1	12.5	8	18.2	9	17.3
	Canada	0	0.0	6	13.6	6	11.5
	Finland	0	0.0	1	2.3	1	1.9
	Germany	0	0.0	1	2.3	1	1.9
	Italy	0	0.0	2	4.5	2	3.8
	Norway	0	0.0	1	2.3	1	1.9
	Portugal	0	0.0	1	2.3	1	1.9
	Sweden	0	0.0	1	2.3	1	1.9
	Korea	0	0.0	1	2.3	1	1.9
	Colombia	0	0.0	1	2.3	1	1.9
Health/disease area	Cancer	0	0.0	2	4.5	2	3.8
	Cardiovascular	0	0.0	4	9.1	4	7.7
	Chronic pain	0	0.0	2	4.5	2	3.8
	Gastrointestinal	0	0.0	3	6.8	3	5.8
	Genitourinary	1	12.5	1	2.3	2	3.8
	High cholesterol	0	0.0	1	2.3	1	1.9
	Musculoskeletal	2	25.0	2	4.5	4	7.7
	Neuropsychiatric	3	37.5	13	29.5	16	30.8
	Non-specific	0	0.0	4	9.1	4	7.7
	Nutrition	0	0.0	1	2.3	1	1.9
	Other	0	0.0	3	6.8	3	5.8
	Physical activity	0	0.0	3	6.8	3	5.8
	Pulmonary/respiratory	1	12.5	4	9.1	5	9.6
	Thyroid disorder	0	0.0	1	2.3	1	1.9
	Diabetes	1	12.5	0	0.0	1	1.9

4.3.2. Sample characteristics

Most of the included studies included adult participants (n=40, 76.9%). 23 (44.2%) included adults aged over 65 years ("older adults") and nine (17.3%) included participants aged under 18 years. The number of patients included in the studies ranged from 1 to 60. The median number of patients was 9. 11 (21.2%) included a single patient only. Of the 12 included study protocols, 10 (83.3%) reported a target sample size; these ranged from 4 to 75 patients.

Table 15. Sample characteristics of n-of-1 studies.

		Disease type					
		Rare		Non-rare		Total	
Variable		*n*	*%*	*n*	*%*	*n*	*%*
Age group	Infant	1	12.5	0	0.0	1	1.9
	Child	0	0.0	8	18.2	8	15.4
	Adult	6	75.0	34	77.3	40	76.9
	Older adult	1	12.5	22	50.0	23	44.2
	Not reported	1	12.5	1	2.3	2	3.8

4.3.3. Intervention and design characteristics

Studies were mostly of pharmaceutical interventions (n=35, 67.3%). Other intervention types were medical devices (n=4, 7.7%), behavioural interventions (n=3, 5.8%), dietary interventions (n=3, 5.8%) and surgery (n=1, 1.9%). The study evaluating a surgical intervention was of deep brain stimulation for essential tremor (33).The remaining 6 studies (11.5%) were classed as 'other' health technologies. They included two studies of blood transfusions (31,32), one investigating methods of tube feeding neonates (34), one of shiatsu (35), one of airway clearance techniques (36) and one looking at the topical application of vegetable oils (37).

The median number of health technologies compared in the studies was two (range 2-4). Studies most commonly compared two health technologies (n=43, 82.7%). Seven studies (13.5%) compared three health technologies, and one compared four (3.8%) (38). The median number of periods was six (range 2-60); 22 studies had six periods (43.3%). The median period length was 14 days (range <1-56). The median number of cycles was three (range 1-16).

Most of the included studies were placebo controlled (n=34, 65.4%). Others used active comparators (n=12, 23.1%), sham devices (n=4, 7.7%), and no intervention arms (n=4, 7.7%). Two studies (3.8%) used more than one control comparator. One used a placebo and a no intervention condition to investigate the side effects of statins (39) and the other used an active comparator and a placebo in a study determining the most suitable placebo for nerve block injections (40).

The majority of studies blinded participants to treatment allocation (n=41, 78.8%). Of those that did not incorporate blinding, three were of behavioural interventions (27.3%) (38,41,42), three were of intervention types classed as 'other' (27.3%) (35,36,43), two were of dietary interventions (18.2%) (34,44), one was of a medical device (9.1%) (45), and one was of a pharmaceutical intervention (9.1%) (46). Blinding was not incorporated either because it was not possible to blind participants due to identifying characteristics of the health technologies or because the patients were infants, and it was therefore assumed to be unnecessary (34).

Washout periods were used in 23 studies (44.3%). The median washout length was 7 days (range 2-90). Washout periods were most commonly two days long (n=5, 9.6%). The median total study duration was 84 days (range 6-406).

Table 16. Intervention and design characteristics of n-of-1 studies

Variable		Disease type					
		Rare		Non-rare		Total	
		n	%	n	%	n	%
Intervention type	Pharmaceutical	6	75.0	29	65.9	35	67.3
	Medical device	1	12.5	3	6.8	4	7.7
	Surgery	0	0.0	1	2.3	1	1.9
	Behavioural	0	0.0	3	6.8	3	5.8
	Other	1	12.5	8	18.2	9	17.3
Comparator type *	Placebo	5	62.5	29	65.9	34	65.4
	Sham	0	0.0	4	9.1	4	7.7
	Active treatment	3	37.5	9	20.5	12	23.1
	No intervention	0	0.0	4	9.1	4	7.7
Blinding employed	Yes	6	75.0	35	79.5	41	78.8
	No	2	25.0	9	20.5	11	21.2
Washout period	Yes	4	50.0	19	43.2	23	44.2
	No	4	50.0	25	56.8	29	55.8

	Rare diseases			Non-rare diseases			Total		
	Median	Min	Max	Median	Min	Max	Median	Min	Max
Number of health technologies	2	2	3	2	2	4	2	2	4
Number of periods	6	2	9	6	2	60	6	2	60
Period length(days)	21	4	28	14	1	56	14	1	56
No. of cycles	3	1	3	3	1	16	3	1	16
Washout length	4.5	2	7	14	2	90	7	2	90
Study length	42	16	252	84	6	406	84	6	406

4.3.4. Outcome characteristics

The majority of studies measured the primary outcome using patient reported outcome measures (PROMs, n=37, 71.2%). Other methods of outcome measurement were using a behavioural test (n=2, 3.8%), a clinical assessment (n=1, 1.9%), a measure of physical activity (n=2, 3.8%), a physiological parameter (n=9, 17.3%). One study (1.9%) included two primary outcomes, of which used a PROM and one which used a lab parameter (47).Most studies measured the primary outcome at regular intervals (n=33, 63.5%). Primary outcomes were most frequently measured daily (n=15, 45.5% of those measuring at regular intervals).

Whether a patient had responded to the treatment was judged in terms of statistical significance in 18 studies (34.6%), clinical significance in 19 studies (36.5%) and not reported in 15 studies (28.8%).

Table 17. Outcome characteristics of n-of-1 trials

Variable		Rare n	Rare %	Non-rare n	Non-rare %	Total n	Total %
Primary outcome measurement	Behavioural test	0	0.0	2	4.5	2	3.8
	Clinical assessment	0	0.0	1	2.3	1	1.9
	Patient report	6	75.0	31	70.5	37	71.2
	Physical activity measure	0	0.0	2	4.5	2	3.8
	Physiological	2	25.0	7	15.9	9	17.3
	Lab parameter + patient report	0	0.0	1	2.3	1	1.9
Outcome measurement at regular intervals	Yes	6	75.0	27	61.4	33	63.5
	No	2	25.0	17	38.6	19	36.5
Frequency of outcome measurement	More than daily	0	0.0	2	4.5	2	3.8
	Daily	2	25.0	13	29.5	15	28.8
	Weekly	3	37.5	6	13.6	9	17.3
	Every 10 days	0	0.0	1	2.3	1	1.9
	Fortnightly	0	0.0	2	4.5	2	3.8
	Four-weekly	1	12.5	2	4.5	3	5.8
	Eight-weekly	0	0.0	1	2.3	1	1.9
	N/a	2	25.0	17	38.6	19	36.5
Definition of treatment response.	Statistical	1	12.5	17	38.6	18	34.6
	Clinical	5	62.5	14	31.8	19	36.5
	Not Given	2	25.0	13	29.5	15	28.8

4.3.5. Analysis characteristics

The included studies used a range of approaches to analysis. The most common of these were t-tests (n=13, 25.0%), regression models (n=14, 26.9%), Bayesian approaches (n=8, 15.4%), and non-parametric analyses (n=8, 15.4%). Other analysis approaches were visual inspection of graphical data (n=2, 3.8%) and time series analysis (n=1, 1.9%). Eight studies stated they did not use formal analysis methods (15.4%) and methods of analysis were not reported in six studies (11.5%).

The majority of studies combined the data from multiple n-of-1 trials or, in the case of study protocols, planned to do so (n=32, 61.5%). There were nine studies which involved multiple participants but did not combine the data across n-of-1 trials (17.3%). Combining data from multiple n-of-1 trials was not possible in trials including only one patient (n=11, 21.2%).

Approximately half of the studies reported numerical results (n=25, 48.1%) Reporting of numerical results was higher in the rare disease studies (n=5, 62.5%) than the non-rare disease studies (n=20, 45.4%). 24 studies (46.1%) reported statistical significance. The significance level was 0.05 in 23 (95.8%) of these and 0.1 in one study (4.2%). P values were reported in 17 studies (32.7%).

Table 18. Analysis characteristics of n-of-1 studies

| | | Disease type | | | | | |
| | | Rare | | Non-rare | | Total | |
Variable		n	%	n	%	n	%
Analysis approach	Bayesian	2	25.0	6	13.6	8	15.4
	Non-parametric	1	12.5	7	15.9	8	15.4
	Graph or visual inspection	0	0.0	2	4.5	2	3.8
	T test	0	0.0	13	29.5	13	25.0
	Regression model	4	50.0	10	22.7	14	26.9
	No formal statistical analysis	0	0.0	8	18.2	8	15.4
	Other	0	0.0	1	2.3	1	1.9
	Not Given	1	12.5	5	11.4	6	11.5
Individual data pooled	Yes	4	50.0	17	38.6	21	40.4
	No	1	12.5	8	18.2	9	17.3
	Planned	2	25.0	9	20.5	11	21.2
	N/a (single patient)	1	12.5	10	22.7	11	21.2
Numerical results reported	Yes	5	62.5	20	45.5	25	48.1
	No	3	37.5	24	54.5	27	51.9
P values reported	Yes	1	12.5	16	36.4	17	32.7
	No	7	87.5	28	63.6	35	67.3

4.3.6. Summary of findings

The review identified reports of 52 n-of-1 trials between January 2011 and May 2021. The characteristics of n-of-1 studies were consistent across those conducted in rare and non-rare disease areas. Studies were conducted in a variety of regions and assessed a wide range of health technologies. 21.2% of studies were of an n-of-1 trial in one patient and the remainder were of a series of n-of-1 trials, including between 2 and 60 participants.

Studies typically assessed drug treatments and were placebo controlled. They had a median of six periods and three cycles. Most studies restricted randomisation with a block size equal to the number of treatment conditions under evaluation. They typically collected PROMs and used statistical significance to determine whether the patient had responded to treatment.

4.4. Conclusion

n-of-1 trials have been implemented in a wide variety of health and disease areas to facilitate evidence-based clinical decision making. Those evaluating health technologies in rare diseases have been designed in a similar way as those in non-rare diseases.

4. Development of key points to consider

5.1. Introduction

In order to develop a set of key points to consider for the design and implementation of n-of-1 trials, a stakeholder workshop, collaborator discussion, and dissemination event were held.

5.2. Stakeholder workshop

A stakeholder workshop was conducted, in which key points to consider for the design and implementation of n-of-1 trials were discussed, in light of the findings from the review.

5.2.1. Aims

- To seek the perspectives of a range of stakeholders on key elements of the design and implementation of n-of-1 trials.
- To generate discussion about the types of questions n-of-1 trials can be used to address, the treatments they can be used to evaluate, and their possible outcomes.

5.2.2. Methods

5.2.2.1. Participants

We aimed to represent the following stakeholders at the workshop:

a) Clinicians who have used n-of-1 trials, or who may wish to use them,
b) statisticians, methodologists, or researchers with an interest in n-of-1 trials,
c) people who have participated in an n-of-1 trial,
d) people living with a rare disease or health condition that requires treatment,
e) others with an interest in n-of-1 trials.

These stakeholder perspectives were considered important for helping us to understand the types of research questions that can be addressed using n-of-1 trials, and when it may or may not be appropriate to undertake n-of-1 trials.

Stakeholders were approached via email, utilising existing contacts within the team and finding additional stakeholders via the following avenues:

a) Online advertisements (e.g. NIHR people in research),
b) Contacting those with published work in this area,
c) Contacting those who could use n-of-1 trials (e.g. GP surgeries, cancer units),
d) Research networks,
e) Word-of-mouth - those invited were welcome to extend the invitation to others they thought might be interested.

Prior to the workshop, participants were provided with an information sheet explaining the purpose and format of the session. Additionally, OH met with the PPI representatives to provide some background information about n-of-1 trials and answer any questions they had, in order that they felt able to engage in the workshop discussions.

5.2.2.2.   Structure of the workshop

The workshop consisted of a presentation of the findings from the review, followed by small group breakout sessions, and then a whole-group feedback session. Due to Covid-19 restrictions, the workshop was held online, via video conferencing, and was recorded to ensure accurate note taking.

Verbal consent was obtained at the start of the workshop and participants were informed that they could leave the workshop at any time.

The following questions were the focus of discussions in the breakout sessions:

- When is it appropriate to use n-of-1 studies? When is it inappropriate?
- When might an n-of-1 be used instead of a parallel group RCT?
- What questions can be addressed using n-of-1 trials?
- What treatments can be assessed using n-of-1 trials?
- What are the benefits and challenges of undertaking and participating in n-of-1 trials?
- What are the considerations of designing n-of-1 trials?

5.2.2.3.   Data management

Video recordings were made of the workshop and breakout discussions. These were stored on the University Google Drive which only the research team had access to and were deleted once notes had been made. Workshop participant personal data (name and contact details) were held in a separate Excel spreadsheet, accessible only by the research team and used only for contact as indicated during consent.

5.2.2.4.   Data analysis

OH listened to the recordings and made detailed notes from each of the discussions, capturing differing perspectives and any examples used to illustrate an idea. The notes were collated and were grouped according to they whether they related to questions, treatments, outcomes, benefits, or challenges.

5.2.3. Results

5.2.3.1.   Participants

The workshop was attended by 13 stakeholders as well as four members of the study team (OH, RC, JR and SAJ). Of the 13, nine were clinicians/researchers, two were statisticians and two were Patient and Public Involvement (PPI) representatives living with a rare disease.

5.2.3.2.   General reflections/benefits and concerns

n-of-1 trials were considered a valuable approach by all workshop attendees. Clinicians/researchers valued the "*perfectly tailored answer"* to the research question that can be obtained by using an n-of-1 trial. Patient representatives also deemed this important, and researchers' experiences of

conducting n-of-1 trials corroborated patient interest in them: *"From my experience… people are very interested in participating [in n-of-1 trials] because it's about them as an individual and whether things work for them or not".* This expectation that the results of an n-of-1 trial will be of direct benefit to the individual patient was considered a benefit of taking part in an n-of-1 trial compared other trial designs.

Workshop participants also felt that a benefit of n-of-1 trials is that the patient will undergo all of the treatment conditions under investigation. This was deemed particularly important in placebo-controlled trials because the patient is guaranteed to undergo the active treatment.

PPI representatives highlighted the fact that patients living with chronic conditions or rare diseases (which n-of-1 trials might investigate) might be particularly attracted to the idea of taking part in an n-of-1 trial, as the formal research process might ensure they "*feel [they are] being taken seriously".*

Clinicians/researchers in the workshop saw difficulties with funding as a key issue with n-of-1 trials. Based on experiences in rare disease research, clinicians anticipated difficultly communicating the justification for trials involving small numbers of patients.

Workshop participants recognised that there might be additional patient burden associated with taking part in an n-of-1 trial. Some sources of burden were not unique to n-of-1 trials, such as attending additional outpatient appointments and questionnaires to fill in, whereas others were enhanced for n-of-1 trials, such as an increased study length compared to parallel group RCTs, as the patient receives multiple treatments over multiple periods and may also receive washout periods in between.

The crossover design was thought to elicit patient burden in another way. If a patient notices improvement under the treatment in one period, there might be psychological difficulty associated with having to stop this treatment in order to undergo another treatment condition. It was understood, however, that having a comparator and multiple treatment periods are an essential part of the trial design and are required in order to provide evidence of an effect in the treatment under investigation. It was agreed that this might be mitigated by the processes of informed consent and, in more challenging circumstances, appropriate psychological/emotional support may mitigate the associated difficulty.

Some clinicians were hesitant about the use of randomisation. This was due to the perception that the patient has less choice and involvement in their treatment, which might be at odds with the idea of personalised medicine and make a patient less likely to engage. Other participants emphasised the importance of randomisation in the n-of-1 design and suggested that the anticipated problems can be mitigated by improving patient understanding of the trial design through the communication about the trial.

5.2.3.3.   Questions
What questions can n-of-1 trials be used to address?

Workshop participants discussed the different types of questions that n-of-1 trials can be used to answer. These are summarised in Table 19. Clinicians in the workshop thought that n-of-1 trials were best suited to comparing active treatments, such as those answering questions about which of the treatment options is preferred by a particular patient. The benefits of trials designed to compare active treatment options is that the patient is undergoing treatment for their condition for the duration of the trial (rather than having placebo periods).

It was also noted that n-of-1 trials are well suited not only to answering questions about treatment efficacy, but also about tolerability and acceptability at the individual level of treatments that have been shown to be efficacious at a group level.

Table 19. Questions that n-of-1 trials can be used to answer

Question	Comparator type	When appropriate?
Does the treatment work at all?	Placebo/sham device	Testing a novel treatment where there is no existing treatment to which it could be compared
Is the treatment better than the existing treatment(s)?	Active treatment	Testing a novel treatment by comparing it to the existing treatment(s)
Which treatment is best for this patient?	Active treatment	Where there are multiple treatment options, but high interindividual variability n-of-1 trials can be used to establish which treatment is most efficacious in an individual patient.
How does the treatment effect vary between individuals?	Active treatment, placebo/sham device, no intervention	When the level of inter individual variability is unknown. Requires a series of n-of-1 trials.

5.2.3.4. Treatments

What treatments can n-of-1 trials be used to investigate?

Participants discussed the treatments that can be investigated using n-of-1 trials. All groups highlighted the requirement that a treatment is reversible in order to be suitable for study using an n-of-1 trial. Reversibility was defined as a patient returning to where they would have been had they not undergone the treatment (i.e. the effect can be washed out once treatment is withdrawn). Participants made a clear distinction between this and a patient returning to baseline. For example, in the case that a disease is slowly progressive, the patient's condition may have worsened following the withdrawal of a reversible treatment, though not as a result of the treatment but rather the progression of the disease. It was agreed that psychological therapies and surgeries are irreversible; their effects cannot be washed out, making them unsuitable treatments to study using n-of-1 trials.

In addition to being reversible, a treatment must have a quick onset and termination of effect in order to be suitable for study using the n-of-1 design. This is in order to avoid the need to incorporate lengthy run in and washout periods into the trial design, which might result in an impractical trial length. One participant provided an example of when an n-of-1 trial may be inappropriate, referencing the deprescribing of steroids following long term use; they have effects that do not disappear quickly after treatment is terminated.

Curative treatments are incompatible with crossover designs, such as n-of-1 trials, because there would be no need to continue to switch between different treatments if the patient's condition has been cured.

There are some treatments which are better suited to assessment using n-of-1 trials than parallel group RCTs. For example, some patients may be ineligible for certain parallel group RCTs on account of having complex or comorbid conditions. These trials often have strict eligibility criteria in order to produce a largely homogenous sample. The sample size in an n-of-1 trial is one and it does not aim to generalise beyond that sample, potentially affording more flexibility with regards to eligibility. Complementary therapies or add on treatments may be well suited for study using n-of-1 trials for this reason.

5.2.3.5. Outcomes

What outcomes should be measured in n-of-1 trials?

Participants noted the importance of having an easily measurable sign or symptom of the condition as an outcome, as well as focusing on outcomes which are most relevant/important to the patient taking part in the study. This might involve the use of personalised outcome measures.

There was discussion regarding the main types of outcomes that can be measured in n-of-1 trials: PROMs, physiological parameters, and biomarkers.

It was suggested that n-of-1 trials should ideally employ both PROMs and a physiological parameter or biomarker. A more objective measure is particularly useful when the patient isn't blinded to treatment allocation, or it is suspected that blinding will be lost due to effects of the treatment.

5.3. Collaborator discussion

5.3.1. Aims
- To reflect on the discussions that were had during the stakeholder workshop.
- To develop key points to consider for n-of-1 study design and implementation.

5.3.2. Methods
A meeting was held between the study team and a collaborator (SAJ, RC, OH and AC) via Google Meets in which draft key points to consider were discussed and developed. SS was unable to attend so his feedback was obtained via email.

5.3.3. Results
The discussion between collaborators raised questions about the language used in the key points to consider. It was agreed that it would be important to define the term 'health technology' to ensure that it was not interpreted as 'gadgets', but rather something that can be applied according to a policy decision.

Much discussion centred around randomisation. The MUSTANG study (48) was discussed as an example of an n-of-1 trial that was not randomised, partly because it would not have been possible to blind the patient to treatment allocation. The relationship between blinding and randomisation was decided to warrant further consideration in revisions of the key points to consider.

It was decided to stipulate in the key points that, where double blinding is not possible, the blinding of some element of outcome assessment should be introduced, e.g. blinded outcome assessor.

The use of more objective outcome measures was agreed to be of additional importance in trials where the patient and/or clinician is unblinded.

The importance of keeping in mind the target audience for these key points was discussed. As they are primarily aimed towards clinicians, rather than academics with methodological expertise, it was decided that it was important to be clear and use limited technical/trial language.

5.4. Dissemination event

5.4.1. Aims

- To share the draft key points to consider with the stakeholders who attended the workshop
- To obtain feedback on the draft key points to consider

5.4.2. Methods

An online dissemination event was held during which draft key points to consider were presented and feedback was requested. Both those that attended the workshop, and those that were invited but were unable to attend, were invited.

5.4.3. Results

Five individuals attended, including a patient representative, a statistician, and three researchers in rare diseases, two of whom had experience of undertaking an n-of-1 trial. Attendees suggested that a key point should be added regarding the need for Patient and Public Involvement (PPI) in n-of-1 trials and suggested that improved links should be made between key points that have an overlapping scope.

Attendees also discussed the scope of the key points to consider; in particular, whether they applied to "informal" n-of-1 trials that may be undertaken outside research settings. Participants discussed that the key points and accompanying report should clarify that they apply only to research undertaken under the supervision of research governance frameworks and where the activity is guided by a research protocol.

5.5. Conclusion

The key points to consider were developed through a series of communications with stakeholders and project collaborators.

6. Key points to consider for n-of-1 trials

6.1. Introduction

Below, a set of key points to consider for the implementation, design, conduct and analysis of n-of-1 trials are provided. Each key point is accompanied by elaboration and case studies where appropriate.

The key points have been divided into two main sections and further sub-divided into additional components, as outlined below:

Section 1: When is it appropriate to undertake n-of-1 trials:

- Scope
- Prevalence of the health condition
- Type and attributes of the health technology
- Questions that can be addressed

Section 2: Design and analysis conditions:

- Choice of outcome
- Choice of comparator
- Target of treatment
- Number of health technologies and periods
- Blinding
- Randomisation
- Analysis

The key points are supplemented by *Table 9,* which presents the questions that can be addressed using n-of-1 trials and the associated design considerations.

6.2. Key points

Table 20. Key points to consider for n-of-1 trials

	Points to Consider	Elaboration
Section 1: When is it appropriate to undertake n-of-1 trials?		
Scope		
1	n-of-1 trials should primarily be used to inform decisions about the care of an individual patient.	An n-of-1 trial should be undertaken where there is a decision to be made regarding the treatment of an individual patient. In some circumstances, an n-of-1 trial could provide sufficient evidence of effect for a health technology to be commissioned for that patient. n-of-1 trials have particular utility where there is large variation in treatment efficacy from patient to patient and so decisions for individual patients are needed.

	Prevalence of health condition	
2	n-of-1 trials can be a viable study design for very low volume interventions, such as those in rare (and ultra-rare) diseases.	There are limited options for the assessment of efficacy for very low volume interventions, such as health technologies for ultra-rare diseases, as the size of the patient population may make it impractical or infeasible to recruit the number of patients required for a conventional parallel group trial. In these cases, n-of-1 trials may be a useful alternative to a conventional parallel group trials as a means of increasing precision when cases are rare. For an example of an n-of-1 trial undertaken in a rare disease, see *case study* 1 (49).
	Type and attributes of health technologies	
3	A wide range of health technologies can be assessed using n-of-1 trials, provided they meet the criteria specified in *points 4 and 5*.	n-of-1 trials can be designed to assess a wide range of health technologies such as drug treatments (see *case study 2* (50)), medical devices (see *case study 3* (51)) and dietary (see *case study 4* (52)) and behavioural interventions (see *case study 5* (53)), provided they meet the criteria specified in points 4 (onset of effect) and 5 (carryover effects).
4	Health technologies to be assessed using n-of-1 trials must have an onset of effect that can feasibly be observed in a study period.	In order to be suitable for study using an n-of-1 trial, a health technology must have an onset of effect that is quick enough that it can be measured within a study period. The onset of effect will impact the length of the periods in an n-of-1 trial and thus the length of the study overall.
5	Health technologies to be assessed using n-of-1 trials must not have prolonged carryover effects.	In an n-of-1 trial, it is important that any carryover effects from one period have expired before an assessment of effect for a subsequent period is conducted. This is to ensure that any effects observed in this assessment can be attributed to the treatment condition of that period. Washout techniques (see point 19) can be employed to ensure that sufficient time has elapsed for carryover effects to expire.
6	n-of-1 trials might be appropriate for the investigation of expensive health technologies or those with significant side effects which effect users to differing extents.	n-of-1 trials can be used to provide evidence of whether the benefits of a health technology outweigh its drawbacks for a particular patient. For an expensive health technology, an n-of-1 trial might be used to assess whether a particular health technology is effective in a particular patient (see *case study 6* (48). If the health technology produces a clinically meaningful effect in a patient, the cost of commissioning it for this patient might be justified. If the health technology is does not produce a clinically meaningful effect in a patient, then the cost of the n-of-1 might be justified by preventing unnecessary costs of commissioning a treatment that does not result in a clinically meaningful improvement for a patient. If a health technology has significant associated side effects which affect users to differing extents, an n-of-1 trial might be used to inform an understanding of the trade-off between benefits and harms for that individual patient. n-of-1 trials are unlikely to be implemented for common, safe, low-cost treatments.

Questions that can be addressed		
7	n-of-1 trials are appropriate when aiming to address one of four questions (see *Table 9*).	The four questions that can be assessed within n-of-1 trials are: ***Does the health technology work at all?*** An n-of-1 trial answering this question will likely be assessing a novel health technology for which there is no evidence in the patient population nor an existing treatment to which it could be compared (see *case study 7* (54)). If the n-of-1 trial was evaluating a drug treatment, the assessment may be of the investigative therapy against placebo (see *case study 8* (55)). ***Does the health technology work better than the existing treatment(s)?*** It might be important to answer this question when there is an existing treatment option for a patient as well as a novel one to be assessed (see *case study 9* (56)). If the n-of-1 trial was evaluating a drug treatment, the assessment may be of the investigative therapy against an active treatment control. ***Which health technology is best for a particular patient?*** This question may be asked in two situations: a) Where a treatment has high patient to patient variability in efficacy - if there is more than one treatment option available with no clear rationale for which will be optimal for a particular patient, for example, because there is high inter-individual variability in treatment response, an n-of-1 trial could be used to determine the treatment choice for each patient (see *case study* 10 (57)). b) Where a number of treatment options are equally efficacious – here a decision will need to be made about on outcomes other than the primary efficacy outcome including factors like patient preference. ***Does the efficacy of the treatment vary between individuals?*** A series of n-of-1 trials would be required to answer this type of question, where each of the individual trials would be answering one of the other questions above. For example, an n-of-1 study could establish the treatment effect of a novel drug treatment compared to placebo for an individual patient. If this were conducted in a number of patients, an assessment could be made of whether the effect is consistent across all patients or within a particular subgroup of patients.

Section 2: Design and analysis considerations for n-of-1 trials		
Choice of outcome		
8	The question being addressed will inform the choice of primary outcome (see *Table 9*).	For most questions, an efficacy outcome will be used as the primary outcome (see *Table 9*).

The choice of efficacy outcome may be influenced by practical considerations such as the time to onset of effect. If the time to onset of effect is long, then an outcome that is usually a secondary outcome (i.e. a surrogate) could be the primary outcome for the study. For example, an early time point assessment of the primary outcome could be used if this is predictive of the final response.

Clinical biomarkers could also be used as the efficacy outcome if these are predictive of the efficacy effect.

If an assessment of efficacy is not the primary research question, then patient preference (see *case study 11* (58)) or quality of life could be the primary outcome. Efficacy outcomes could be secondary outcomes in such a study. |
| 9 | It is recommended to use both patient reported outcome measures (PROMs) and more objective measures of effect where possible, especially in those trials that are being undertaken to assess the efficacy of an expensive or risky treatment. | Those trials that are being undertaken to assess the efficacy of an expensive or risky treatment may require more stringent design considerations than those trials that are being undertaken on a more informal basis to dictate care (see also *blinding – point 20*). An example of this is *case study 12* (48).

In such a situation, sufficient evidence of clinical improvement is required. For example, just collecting patient preference may not be sufficient, but more objective measures of a clinically significant effect may be required, as well as patient reported outcome measures (PROMs). |
| 10 | n-of-1 trials can be used not only to assess the effect of a health technology on a primary efficacy outcome but also other outcomes which are important to the patient. | n-of-1 trials can be used not only to assess the effect of a health technology on the primary efficacy outcome, but also other outcomes which are important to the patient. Such outcomes could be the primary outcome, or secondary outcomes, for the trial. For an example, see *case study 13* (59).

For example, an n-of-1 trial could be used to assess the effect of different therapies on treatment side effects.

Alternatively, patient preference for care delivery might be assessed.

n-of-1 trials might make the personalisation of outcomes possible. |
| **Choice of comparator** | | |
| 11 | The choice of comparator should be made to answer the research question for the study. | Different comparators are appropriate to answer different research questions. See elaboration of point 7 and Table 21. |

	Target of treatment	
12	n-of-1 trials are used to provide evidence which can be used to improve the patient's condition itself, specific symptoms of the condition, side effects, or patient satisfaction.	An n-of-1 study can be undertaken to assess a health technology in the improvement of the: • Condition/disease itself – the patient will benefit as their health condition will improve (see *case study 14* (60)); • Symptoms of the condition – the patient will benefit as their quality of life will improve (see *case study 15* (61)); • Side effects - the patient will benefit as their quality of life will improve but also their health condition may improve as the treatment may be better tolerated, improving adherence (see case study 16 (62)); • Patient satisfaction – if two health technologies have equal efficacy but with different posologies patient preference could determine the choice of treatment. The patient will benefit as they get the treatment which works best for them in terms of their daily life.
	Number of health technologies and periods	
13	n-of-1 trials typically compare two health technologies. Designing n-of-1 trials which compare three or more health technologies is associated with practical challenges.	Most n-of-1 trials compare two health technologies (e.g. drug and placebo or two active treatments – see *case study 17* (51)). Designing these trials is more straightforward than those evaluating more than two health technologies, which incur greater practical and logistical challenges such an increased study duration (see *case study 18* (56)). It is possible to conduct n-of-1 trials evaluating more than three health technologies, particularly if the period and washout lengths are short (*see case study 19* (63)), however it might be preferable to instead conduct more than n-of-1 trial.
14	The number of study periods in an n-of-1 trial is a trade-off between precision and feasibility.	The more study periods there are in an n-of-1 trial, the greater the precision in the evaluation of effect as there are more evaluations of the health technologies. Decisions about the number of study periods must take into consideration the overall study length. For some n-of-1 trials, having many study periods may be practicable. For others it both might not be practical or even required - for a patient preference study it might be possible to get an answer in just two study periods. The DIAMOND review of n-of-1 studies found that the median number of study periods in an n-of-1 trial was six (see *case study 20* (64)). This seems to represent a balance between precision and feasibility.

Blinding

15	n-of-1 trials should be blinded where feasible.	Where feasible, n-of-1 trials should be double blinded – for an example, see *case study 21* (65). Blinding is more difficult in n-of-1 trials of certain types of health technologies, such as behavioural or dietary interventions.
		Blinding may be challenging in n-of-1 trials of drug treatments due to difficulty obtaining a suitable placebo or due to obvious differences in the appearance or effects of active treatments to be compared.
		Blinding may be more important in n-of-1 trials because of crossover design.
		If a double-blind is not possible, n-of-1 trials may be conducted as single-blind or open label trials.
		For open label trials it is optimal to incorporate some blinding, such as blind assessment of outcomes.

Randomisation

16	Blocked randomisation of treatment allocation is typically recommended.	Randomising the sequence of treatment allocation has the advantage of evenly distributing (on average) both known and unknown confounding factors between the health technologies.
		Blocking randomisation using a block size equal to the number of health technologies in the study has the advantage of preventing the generation of undesirable sequences such as AAAABBBB which would make the study sensitive to drop out as, if a patient dropped out halfway through the study, they would only have data from treatment condition A that could be analyses. A block size of two will produce a sequence such as ABBAABAB.
		In this example, if a patient dropped out of the study after two periods, it would still be possible to assess their experience of both health technologies for the patient.
		Even if there is no patient dropout there could be issues with an allocation AAAABBBBB if there is a time effect for the underlying condition in the patient. A sequence of the form ABBAABAB would help to mitigate against this.
		A pitfall of randomising with a small block size is that the patient and clinicians are more likely to work out or guess the treatment allocation.
		It is typically not possible to conceal the block size from a patient in an n-of-1 trial, in order to uphold the ethical and legal requirement of informed consent.
		The risk of working out or guessing the treatment allocation should be weighed against the risk of patient withdrawal.

	Analysis	
18	An interim analysis may be considered when designing n-of-1 trials. The analysis can be used to indicate whether early stopping of the trial is appropriate.	An interim analysis may be considered if there are six or more study periods and the study is long enough to assess effect in the reduced number of periods. Continuing switching between treatments may become difficult if the patient is experiencing a noticeable improvement or deterioration in their symptoms under one particular treatment such that there is sufficient evidence of effect for an individual patient prior to the planned completion of the trial. The rationale for stopping the study early should be done on a study-by-study basis and where possible should be pre-specified. See point 22 as information from other patients may inform the decision.
19	Washout periods or active (analytical) washout should be employed if there are likely to be carryover effects of the health technology under investigation.	Where there are likely to be carryover effects from one study period to the next, a washout period should be implemented between study periods – for example, see *case study 22* (66)). The required length of washout will be determined by characteristics of the intervention. Where it would be inappropriate to withdraw treatment for a period of washout, the use of active washout should be considered. The strategy of an *active washout* can be used, whereby patients can be switched immediately from one treatment to another (if safe to do so) but measurement starts once the effect of the previous treatment has disappeared and steady state has been reached. Although longer washout periods are generally desirable it can potentially lead to harm for the patients if treatment was withdrawn and therefore full washout can raise ethical concerns. An active washout therefore is valid design when full washout will lead to harm. With the design the assumption is when we make the clinical assessments the efficacy will be for the treatment in that pathway. It is worth noting that carry-over is not just influenced by treatment. For outcomes such as patient reported outcomes there can be psychometric carry-over as patients can recall how they responded in previous periods.
20	Clinical, in addition to statistical, significance should be used to help judge the effect of treatment.	Determining the effect of treatment based solely on statistical significance should be avoided. Clinical significance should also be considered (see *case study 23* (62)). Where possible, a definition of a clinically important effect should be defined in the protocol, in order to introduce a degree of objectivity to an otherwise subjective assessment.
21	Within-patient analysis of an n-of-1 trial will determine whether a clinically important effect has been observed.	Within-patient analysis of an n-of-1 trial will determine whether a clinically important effect has been observed. Replication is informative in the assessment of response as it enables an assessment of the personalised response to treatment for an individual patient including if the effect is consistent or varies.

22	Between-patient analysis of a series of n-of-1 trials can be used to estimate the average treatment effect across all the trials; determine whether these effects are consistent for all of the patients and to estimate the average treatment effect for that population/sub-population generally.	If a series of n-of-1 trials is conducted in which there is a consistent effect of treatment observed across all patients (or in a subgroup of patients), then it would be possible to combine the individual estimates of effect using meta-analysis to obtain an overall estimate of effect. See *case study 24* (56). Quantifying the effect within individual patients is still the primary analysis (see principle 2), but a meta-analysis is informative. These estimates will inform clinical practice overall – for example a recommendation could be made for all patients to receive the new treatment including those who have been in an n-of-1- and also clinical decisions for the individual patients in the meta-analysis – as if the effects are consistent outcomes seen overall can be used for individual patients.
PPI		
23	Relevant and meaningful PPI should be sought throughout the n-of-1 trial including design and planning; interpretation; dissemination and implementation.	PPI might be especially important in n-of-1 trials due to their personalised nature. Input may be sought from the patient who will be taking part, disease specific charities, affiliated support groups or hospital/Trust specific advisory groups. During the design and planning of the trial, input may be sought into the patient facing materials (e.g., PIS), outcomes and the treatment and follow-up regimes. During interpretation and dissemination, input may be sought into how the results are presented and shared with other patients.

Table 21. Design considerations for n-of-1 trials

Question [a]	Does the health technology work at all for a particular patient?	Does the health technology work better than the existing treatment(s) for a particular patient?	Which health technology is best for a particular patient?		Does the individual treatment effect vary between patients?
			When there is high variability in effect between patients	When there is a number of equally efficacious treatments	
Design	Individual n-of-1 trial	Individual n-of-1 trial	Individual n-of-1 trial	Individual n-of-1 trial	Series of n-of-1 trials
Primary outcome [b]	Efficacy	Efficacy	Efficacy	Patient preference	Efficacy
Control [c]	Placebo (drug trial) / standard of care (behavioural or other trial)	Active treatment	Active treatment	Active treatment	Placebo / standard of care / active treatment

[a] See point 7

[b] see points 8, 9 and 10

[c] see point 11

7. Discussion

The DIAMOND project has developed twenty-one key points for researchers and clinicians to consider when designing and undertaking n-of-1 trials, focussing on when it is appropriate to undertake n-of-1 trials, the questions that can be addressed within such trials, and design and analysis considerations. In order to develop these key points we firstly updated a review by Gabler et al. (2011) in order to identify n-of-1 trials and their characteristics (7). In our updated review we identified 52 trials, eight of which were undertaken in rare diseases. The studies most frequently compared drug treatments and used a placebo control. The most frequently used number of periods was six and number of cycles was three. We did not identify any design differences between n-of-1 trials in rare and non-rare diseases. The results of the review informed the key points for n-of-1 trials, developed through consultation of key stakeholders via workshops, email contact and meetings. We have also developed freely available resources to assist others in undertaking these trials – namely, statistical packages for the analysis of such trials and training courses for both statisticians and non-statisticians (67).

Strengths and limitations

This study involved the thorough review of previously undertaken n-of-1 trials from 2011 until 2021, and therefore represents an up-to-date audit of randomised n-of-1 trials that have been conducted over this time period. Additionally, it is the first review to compare the design characteristics of n-of-1 trials conducted in rare and non-rare diseases. The review informed a workshop involving 13 stakeholders in the field of n-of-1 trials from a range of backgrounds and disciplines as well as two patient representatives. In order to ensure they were generalisable, the final list of key points for n-of-1 trials was also reviewed by the stakeholders at a dissemination event. Limitations include that, with only 13 stakeholders, the results may not be generalisable compared to if a larger group of individuals were assembled.

Comparison to other studies

The DIAMOND study builds on other similar studies in the area of n-of-1 trials. Firstly, a detailed report by Kravitz et al. (2014) also summaries key considerations for designing and conducting n-of-1 trials (19). The Kravitz et al. report (19) was commissioned by the US government and is therefore US focussed and aimed at a wide audience (including patients, statisticians, researchers). In comparison, the DIAMOND study engaged with predominantly UK-based stakeholders to develop UK focussed key points for n-of-1 trials targeted specifically at clinicians or trials methodologists who are interested in

undertaking such trials. The DIAMOND study builds on the work undertaken by Kravitz et al. by updating a review of recently published trials, and a workshop with key stakeholders, which has resulted in additional key points being developed, including a list of questions that can be addressed by n-of-1 trials, the consideration of an interim analysis part way through an n-of-1 trial, and how the cost of a new treatment may influence the design of the trial. We have also developed key materials, including software and training packages, for researchers in this area to enable them to design and conduct such trials, and to minimise any barriers to undertaking n-of-1 trials.

Secondly, our study updates the systematic review undertaken by Gabler et al. (7), which also aimed to summarise the characteristics of previously conducted n-of-1 trials. Our review identified a further 52 n-of-1 trials undertaken since the original Gabler et al. (2011) review. The characteristics of those identified in our review broadly match those identified by Gabler et al. However, there are a few exceptions. Firstly, our review identified a higher number of studies that evaluated behavioural interventions (Gabler et al n=1, 1%, our review n=3, 5.8%) -this might either reflect a genuine increase in the number of these trials or result from the different search strategies used, with Gabler et al searching only medical journals, whereas we did not put any restrictions on the type of journal. Secondly, our review identified a lower occurrence of blinding (Gabler et al n=106, 98%, our review n=41, 78.8%), which is likely to be a consequence of the differing proportions of intervention types identified, i.e. more behavioural interventions and those classed as 'other', as these are typically more difficult to blind than trials of pharmaceutical interventions. Finally, there were some differences in the approaches to analysis. Our review identified a lower rate of analysis by visual inspection of graphs (Gabler et al n=56 52% our review n=2, 3.8%) and a higher rate of pooled analysis than that of Gabler et al (Gabler et al n=26, 24%, our review n=21, 40.4%). These differences may be due to the changing research landscape over time, with pressures to report p-values and to undertake studies that can estimate population treatment effects.

Barriers remain that may hinder clinicians and researchers in undertaking n-of-1 trials. One such obstacle is likely to be seeking regulatory (e.g. ethical) approvals for n-of-1 trials, which has been reported by several authors (68–70). There has been discussion in the literature regarding whether n-of-1 trials require review by an ethics committee, stemming from a debate regarding whether such studies are medical research or an optimised form of clinical care (69,70). Several papers, based on the regulatory contexts in the USA and Netherlands, have looked to clarify when ethical approvals are required for n-of-1 trials, with the inclusion of only one participant (i.e. a single n-of-1 trial) being an important exclusionary factor to requiring approvals (68,70). According to one paper, drugs approved

for the indication in which it will be used, or an off-label drug with mild side-effects, with no extra invasive or burdensome measures may not require ethical approvals (68). In the UK setting, according to the MHRA algorithm, a study that involves allocation of treatments decided in advance by a trial protocol requires MHRA regulatory approvals (71).

There are some methodological considerations discussed in the literature that are not covered by the DIAMOND work, such as non-randomised sequence allocations. Randomising the allocation of treatment to period is an important technique to reduce bias in an n-of-1 trial, so much so that some consider randomisation a defining feature of an n-of-1 trial (10,72). Some authors, however, suggest an alternative approach under particular circumstances. In their 2014 report, Kravitz et al. (19) suggest that counterbalancing (generation of a balanced, non-randomised sequence of treatment periods e.g. AB BA BA AB) is an appropriate alternative to randomisation when there are known time trends of the condition being studied (i.e. it is deteriorating). Whilst we do not suggest counterbalancing as an alternative approach, our key points recommend blocking the randomisation by the number of health technologies under investigation. This goes some way to balancing the sequence to control for a known effect of time on the patient's condition, because a sequence such as AAABBB cannot be produced. Additionally, if time trends are a key concern, then a crossover design will not be an optimal method of evaluation of a treatment as they rely on the condition being stable throughout the evaluation.

Implications

The key points provide guidance that may assist clinicians who are interested in an n-of-1 study to decide whether it is an appropriate choice. If it is an appropriate choice, the key points may support the clinician with the design and implementation of the study.

The two courses on the design and analysis of n-of-1 trials were well attended. This has resulted in more people being aware of the n-of-1 design and the statistical analysis of such studies. The additional resources we have made available on the website (67) will enable people to conduct the analyses explained in the courses.

The use of the outputs of the project will potentially result in more thoughtfully designed n-of-1 studies, which would result in a better use of research resources and improved outcomes for patients.

Recommendations for future research

The outputs from the DIAMOND study could be further developed into a tool that allows for the development of a tailored protocol for a particular type of n-of-1 study, therefore allowing researchers and clinicians to implement the design and DIAMOND key points for n-of-1 trials.

References

1. Mirza R, Punja S, Vohra S, Guyatt G. The history and development of N-of-1 trials. J R Soc Med. 2017 Aug 4;110(8):330–40.

2. Selker HP, Cohen T, D'Agostino RB, Dere WH, Ghaemi SN, Honig PK, et al. A Useful and Sustainable Role for N-of-1 Trials in the Healthcare Ecosystem. Clin Pharmacol Ther. 2022;112(2):224–32.

3. Kravitz RL, Paterniti DA, Hay MC, Subramanian S, Dean DE, Weisner T, et al. Marketing therapeutic precision: potential facilitators and barriers to adoption of n-of-1 trials. Contemp Clin Trials. 2009;30(5):436–45.

4. Gupta S, Faughnan ME, Tomlinson GA, Bayoumi AM. A framework for applying unfamiliar trial designs in studies of rare diseases. J Clin Epidemiol. 2011 Oct;64(10):1085–94.

5. Nikles J, Mitchell G. Introduction. In: The Essential Guide to N-of-1 Trials in Health. Dordrecht: Springer Netherlands; 2015. p. 1–7.

6. 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;86(4):533–55.

7. Gabler NB, Duan N, Vohra S, Kravitz RL. N-of-1 trials in the medical literature: a systematic review. Med Care. 2011 Aug;49(8):761–8.

8. Vandenbroucke JP. The N-of-1 trial: the ideal study design that is underused. Ned Tijdschr Geneeskd. 2006;150(51):2794–5.

9. Sibbald B, Roland M. Understanding controlled trials. Why are randomised controlled trials important? BMJ. 1998;316(7126):201.

10. Guyatt G, Sackett D, Taylor DW, Ghong J, Roberts R, Pugsley S. Determining optimal therapy—randomized trials in individual patients. N Engl J Med. 1986;314(14):889–92.

11. Rare diseases [Internet]. [cited 2022 Nov 1]. Available from: https://research-and-innovation.ec.europa.eu/research-area/health/rare-diseases_en

12. Abrahamyan L, Feldman BM, Tomlinson G, Faughnan ME, Johnson SR, Diamond IR, et al. Alternative designs for clinical trials in rare diseases. In: American Journal of Medical Genetics Part C: Seminars in Medical Genetics. Wiley Online Library; 2016. p. 313–31.

13. The Lancet Neurology. Rare neurological diseases: a united approach is needed. Lancet Neurol. 2011 Feb 1;10(2):109.

14. Ferreira CR. The burden of rare diseases. Am J Med Genet A. 2019;179(6):885–92.

15. Punja S, Bukutu C, Shamseer L, Sampson M, Hartling L, Urichuk L, et al. N-of-1 trials are a tapestry of heterogeneity. J Clin Epidemiol. 2016;76:47–56.

16. Vohra S, Shamseer L, Sampson M, Bukutu C, Schmid CH, Tate R, et al. CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement. BMJ. 2015 May 14;350:h1738.

17.   Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. J Pharmacol Pharmacother. 2010;1(2):100–7.

18.   Müller AR, Brands MM, van de Ven PM, Roes KC, Cornel MC, van Karnebeek CD, et al. Systematic review of N-of-1 studies in rare genetic neurodevelopmental disorders: the power of 1. Neurology. 2021;96(11):529–40.

19.   Kravitz RL. and the DEcIDE Methods Center N-of-1 Guidance Panel (Duan N, Eslick I, Gabler NB, Kaplan HC, Kravitz RL, Larson EB, Pace WD, Schmid CH, Sim I, Vohra S). Des Implement N--1 Trials User's Guide AHRQ Publ. 2016;(13):14.

20.   Chalmers I, Smeeth L, Goldacre B. Personalised medicine using N-of-1 trials: overcoming barriers to delivery. In: Healthcare. MDPI; 2019. p. 134.

21.   Jaeschke R, Adachi J, Guyatt G, Keller J, Wong B. Clinical usefulness of amitriptyline in fibromyalgia: the results of 23 N-of-1 randomized controlled trials. J Rheumatol. 1991 Mar 1;18(3):447–51.

22.   Zucker DR, Schmid CH, McIntosh MW, D'Agostino RB, Selker HP, Lau J. Combining single patient (N-of-1) trials to estimate population treatment effects and to evaluate individual patient responses to treatment. J Clin Epidemiol. 1997 Apr 1;50(4):401–10.

23.   Zucker DR, Ruthazer R, Schmid CH. Individual (N-of-1) trials can be combined to give population comparative treatment effect estimates: methodologic considerations. J Clin Epidemiol. 2010 Dec 1;63(12):1312–23.

24.   Understanding Variation in Sets of N-of-1 Trials | PLOS ONE [Internet]. [cited 2022 Oct 11]. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0167167

25.   Senn S. Controversies concerning randomization and additivity in clinical trials. Stat Med. 2004;23(24):3729–53.

26.   Araujo A, Julious S, Senn S. Understanding variation in sets of N-of-1 trials. PloS One. 2016;11(12):e0167167.

27.   Senn SJ. in defence of the linear model. Control Clin Trials. 2000;21:589–92.

28.   Schwarzer G. meta: An R package for meta-analysis. R News. 2007;7(3):40–5.

29.   Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat Softw. 2010;36(3):1–48.

30.   Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021 Mar 29;372:n71.

31.   Seitelbach M, Chin-Yee IH, Kinney J, Chia J, Ormond K, Mahon J, et al. Age of Blood Does Not Affect Quality of Life and Hemoglobin: N–of- 1 Trials in Transfusion-Dependent Patients. Blood. 2011;118(21):3369.

32.   Hsia CC, Mahon JL, Seitelbach M, Chia J, Zou G, Chin-Yee IH. Use of n-of-1 (single patient) trials to assess the effect of age of transfused blood on health-related quality of life in

transfusion-dependent patients: AGE OF BLOOD N-OF-1 TRIALS. Transfusion (Paris). 2016 May;56(5):1192–200.

33.   Hyam JA, Pereira EAC, McCulloch P, Javed S, Plaha P, Mooney L, et al. Implementing novel trial methods to evaluate surgery for essential tremor. Br J Neurosurg. 2015 May 4;29(3):334–9.

34.   Jensen EA, Zhang H, Feng R, Dysart K, Nilan K, Munson DA, et al. Individualising care in severe bronchopulmonary dysplasia: a series of N-of-1 trials comparing transpyloric and gastric feeding. Arch Dis Child - Fetal Neonatal Ed. 2020 Jul;105(4):399–404.

35.   Tsiormpatzis S. Effects of Shiatsu on the Health-Related Quality of Life of a Person with Secondary Progressive Multiple Sclerosis. 2019;

36.   Gursli S, Sandvik L, Bakkeheim E, Skrede B, Stuge B. Evaluation of a novel technique in airway clearance therapy – Specific Cough Technique (SCT) in cystic fibrosis: A pilot study of a series of N-of-1 randomised controlled trials. SAGE Open Med. 2017 Dec;5:205031211769750.

37.   Santos C, Weaver DF. Topically applied linoleic/linolenic acid for chronic migraine. J Clin Neurosci. 2018;58:200–1.

38.   Brannon EE, Cushing CC, Walters RW, Crick C, Noser AE, Mullins LL. Goal feedback from whom? A physical activity intervention using an N-of-1 RCT. Psychol Health. 2018;33(6):701–12.

39.   Howard JP, Wood F, Finegold J, Nowbar A, Thompson D, Arnold AD, et al. A Three-arm N-of-1 Trial With Statin, Placebo and Tablet Free Periods, to Verify Side Effects and Identify Their Cause: The SAMSON Trial. In: CIRCULATION. LIPPINCOTT WILLIAMS & WILKINS TWO COMMERCE SQ, 2001 MARKET ST, PHILADELPHIA …; 2020. p. E480–1.

40.   Smith JH, Potter JL, Robblee JV. What Is the Optimal Placebo for Clinical Trials of Nerve Blocks in Headache?: An N-of-1 Study of Lactated Ringer's. Headache J Head Face Pain. 2019;59(1):108–9.

41.   Cushing C. N-OF-1 RCT FOR DEVELOPING A CONTROL THEORY SMS INTERVENTION FOR ADOLESCENT PHYSICAL ACTIVITY. In: ANNALS OF BEHAVIORAL MEDICINE. SPRINGER 233 SPRING ST, NEW YORK, NY 10013 USA; 2015. p. S50–S50.

42.   Sniehotta FF, Presseau J, Hobbs N, Araújo-Soares V. Testing self-regulation interventions to increase walking using factorial randomized N-of-1 trials. Health Psychol. 2012;31(6):733.

43.   Tian Y, Ma Y, Fu Y, Zheng JS. Application of n-of-1 clinical trials in personalized nutrition research: a trial protocol for Westlake N-of-1 Trials for Macronutrient Intake (WE-MACNUTR). Curr Dev Nutr. 2020;4(9):nzaa143.

44.   Kaplan HC, Opipari-Arrigan L, Schmid CH, Schuler CL, Saeed S, Braly KL, et al. Evaluating the comparative effectiveness of two diets in pediatric inflammatory bowel disease: a study protocol for a series of N-of-1 trials. In: Healthcare. MDPI; 2019. p. 129.

45.   Benhamou PY, Lablanche S, Vambergue A, Doron M, Franc S, Charpentier G. Patients with highly unstable type 1 diabetes eligible for islet transplantation can be managed with a closed-loop insulin delivery system: a series of N-of-1 randomized controlled trials. Diabetes Obes Metab. 2021;23(1):186–94.

46.   Samuel JP, Tyson JE, Green C, Bell CS, Pedroza C, Molony D, et al. Treating hypertension in children with n-of-1 trials. Pediatrics. 2019;143(4).

47.   Sun M, Chai L, Lu F, Zhao Y, Li Q, Cui B, et al. Efficacy and Safety of Ginkgo Biloba Pills for Coronary Heart Disease with Impaired Glucose Regulation: Study Protocol for a Series of *N* -of-1 Randomized, Double-Blind, Placebo-Controlled Trials. Evid Based Complement Alternat Med. 2018 Oct 14;2018:1–8.

48.   ISRCTN - ISRCTN17945917: Investigating the role of miglustat in the management of a patient with Tangier Disease [Internet]. [cited 2022 Dec 13]. Available from: https://www.isrctn.com/ISRCTN17945917

49.   Benhamou P, Lablanche S, Vambergue A, Doron M, Franc S, Charpentier G. Patients with highly unstable type 1 diabetes eligible for islet transplantation can be managed with a closed-loop insulin delivery system: A series of N-of-1 randomized controlled trials. Diabetes Obes Metab. 2021 Jan;23(1):186–94.

50.   Roustit M, Giai J, Gaget O, Khouri C, Mouhib M, Lotito A, et al. On-Demand Sildenafil as a Treatment for Raynaud Phenomenon. Ann Intern Med. 2018 Nov;169(10):694.

51.   Frost KL, Carey JR, Broback TW, Carlson NL, Daggett CA, Dalbec MM, et al. N-of-1 Trial in Person with Pontine Stroke Receiving Repetitive Transcranial Magnetic Stimulation to Improve Hand Function. J Neuroimaging Psychiatry Neurol. 2017;2(2):36–42.

52.   Tian Y, Ma Y, Fu Y, Zheng JS. Application of n-of-1 Clinical Trials in Personalized Nutrition Research: A Trial Protocol for Westlake N-of-1 Trials for Macronutrient Intake (WE-MACNUTR). Curr Dev Nutr. 2020 Sep;4(9):nzaa143.

53.   Rvachew S, Matthews T. An N-of-1 Randomized Controlled Trial of Interventions for Children With Inconsistent Speech Sound Errors. J Speech Lang Hear Res JSLHR. 2019;62(9):3183–203.

54.   McGarry ME, Illek B, Ly NP, Zlock L, Olshansky S, Moreno C, et al. In vivo and in vitro ivacaftor response in cystic fibrosis patients with residual CFTR function: N-of-1 studies. Pediatr Pulmonol. 2017;52(4):472–9.

55.   Tsiormpatzis S. Effects of shiatsu on the health-related quality of life of a person with secondary progressive multiple sclerosis: A mixed methods N-of-1 trial within a whole systems research case study. Eur J Integr Med. 2019 Dec;32:101006.

56.   Samuel JP, Tyson JE, Green C, Bell CS, Pedroza C, Molony D, et al. Treating Hypertension in Children With n -of-1 Trials. Pediatrics. 2019 Apr;143(4):e20181818.

57.   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. 2018 Dec 11;320(22):2344.

58.   Smith JH, Potter JL, Robblee J V. What Is the Optimal Placebo for Clinical Trials of Nerve Blocks in Headache?: An N-of-1 Study of Lactated Ringer's. Headache J Head Face Pain. 2019 Jan;59(1):108–9.

59.   Graham T, Sooriah S, Box R, Gage H, Williams P, Clemett V, et al. Participatory co-design of patient-reported outcome indicators and N-of-1 evaluation of a dressing glove for Epidermolysis bullosa. J Wound Care. 2020 Dec 2;29(12):751–62.

60.   Santos C, Weaver DF. Topically applied linoleic/linolenic acid for chronic migraine. J Clin Neurosci. 2018 Dec;58:200–1.

61.   Ferreira JJ, Mestre T, Guedes LC, Coelho M, Rosa MM, Santos AT, et al. Espresso Coffee for the Treatment of Somnolence in Parkinson's Disease: Results of n-of-1 Trials. Front Neurol. 2016 Mar 8;7.

62.   Joy TR, Monjed A, Zou GY, Hegele RA, McDonald CG, Mahon JL. N-of-1 (single-patient) trials for statin-related myalgia. Ann Intern Med. 2014 Mar;160(5):301–10.

63.   Brannon EE, Cushing CC, Walters RW, Crick C, Noser AE, Mullins LL. Goal feedback from whom? A physical activity intervention using an N-of-1 RCT. Psychol Health. 2018 Jun 3;33(6):701–12.

64.   Lipka AF, Vrinten C, van Zwet EW, Schimmel KJM, Cornel MC, Kuijpers MR, et al. Ephedrine treatment for autoimmune myasthenia gravis. Neuromuscul Disord. 2017 Mar;27(3):259–65.

65.   Cha YH, Gleghorn D, Doudican BC. Double-blind randomized N-of-1 trial of transcranial alternating current stimulation for mal de débarquement syndrome. Carvalho S, editor. PLOS ONE. 2022 Feb 4;17(2):e0263558.

66.   Germini F, Coerezza A, Andreinetti L, Nobili A, Rossi PD, Mari D, et al. N-of-1 Randomized Trials of Ultra-Micronized Palmitoylethanolamide in Older Patients with Chronic Pain. Drugs Aging. 2017 Dec 5;34(12):941–52.

67.   University of Sheffield. Development of generalisable methodology for n-of-1 trials delivery for very low volume treatments (DIAMOND) [Internet]. [cited 2022 Dec 13]. Available from: https://www.sheffield.ac.uk/scharr/research/centres/ctru/diamond

68.   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 Feb 27;8(1):49.

69.   McDonald S, Nikles J. N-of-1 Trials in Healthcare. Healthcare. 2021 Mar 15;9(3):330.

70.   Cen R, Hussain A, Pak KJ, Mitchell G, Nikles J, Gaudreau S, et al. Do N-of-1 Trials Need IRB Review? J Empir Res Hum Res Ethics. 2016 Jul;11(3):250–5.

71.   Is it a clinical trial of a medicinal product? Algorithm_Clean__1_.pdf [Internet]. [cited 2022 Dec 13]. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/949145/Algorithm_Clean__1_.pdf

72.   Margolis A, Giuliano C. Making the switch: From case studies to N-of-1 trials. Epilepsy Behav Rep. 2019 Jan 1;12:100336.

Appendix 1

Variables for extraction

First Author

Year of publication

Condition under study

Health/disease area

Rare disease (Yes/No)

Funding reported (Yes/no)

Region (If unclear, this was taken to be the region of the first author's institution)

Protocol (Yes/No)

Target sample size (For protocols only)

Intervention type (pharmaceutical, behavioural, dietary, medical device, other)

Data pooled? (Yes/No/Planned/Not applicable – single patient)

Number of patients included

Number of health technologies compared (includes control groups)

Comparison type (placebo, sham device, no intervention arm, active control)

Infants included – where this was not specified in the eligibility criteria, the characteristics of the included participants were used as a proxy

Children included – where this was not specified in the eligibility criteria, the characteristics of the included participants were used as a proxy

Adults included – where this was not specified in the eligibility criteria, the characteristics of the included participants were used as a proxy

Older adults included – where this was not specified in the eligibility criteria, the characteristics of the included participants were used as a proxy

Number of periods

Period length (days)

Total study duration (days)

Washout period (Yes/No)

Washout length

Primary outcome type (Patient report, physiological, behavioural test, physical activity, clinical assessment, lab parameter)

Outcome measurement at regular intervals (Yes/No)

Frequency of outcome measurement

Blinding (Yes/No)

Definition of response to treatment (Statistical, clinical)

Analysis approach

Numerical results (Yes/No)

Standard deviation given (Yes/no)

P values given (Yes/No)

Significance level used

Appendix 2

References of studies included in the review

Bech AP, Wetzels JF, Groenewoud H, Nijenhuis T. The use of N-of-1 trials to individualize treatment in patients with renal magnesium wasting. American Journal of Kidney Diseases. 2019 Feb 1;73(2):288-90.

Benhamou PY, Lablanche S, Vambergue A, Doron M, Franc S, Charpentier G. Patients with highly unstable type 1 diabetes eligible for islet transplantation can be managed with a closed-loop insulin delivery system: a series of N-of-1 randomized controlled trials. Diabetes, Obesity and Metabolism. 2021 Jan;23(1):186-94.

Brannon EE, Cushing CC, Walters RW, Crick C, Noser AE, Mullins LL. Goal feedback from whom? A physical activity intervention using an N-of-1 RCT. Psychology & health. 2018 Jun 3;33(6):701-12.

Ennis JD, Chari VR, Graham A, Harvey DT, Ho E, Nesathurai S. Poster 66 Levodopa/Carbidopa to Improve Motor Recovery After Brain Tumor Excision: An N-of-1 Case Report. PM&R. 2012 Oct;4:S211.

Chen YM, Deng JM, Wen Y, Chen B, Hou JT, Peng B, Zhang SJ, Mi H, Jiang QL, Wu XL, Liu FB. Modified Sijunzi decoction in the treatment of ulcerative colitis in the remission phase: study protocol for a series of N-of-1 double-blind, randomised controlled trials. Trials. 2020 Dec;21(1):1-7.

Cushing C. N-OF-1 RCT FOR DEVELOPING A CONTROL THEORY SMS INTERVENTION FOR ADOLESCENT PHYSICAL ACTIVITY. InANNALS OF BEHAVIORAL MEDICINE 2015 Apr 1 (Vol. 49, pp. S50-S50). 233 SPRING ST, NEW YORK, NY 10013 USA: SPRINGER.

Ferreira JJ, Mestre T, Guedes LC, Coelho M, Rosa MM, Santos AT, Barra M, Sampaio C, Rascol O. Espresso Coffee for the Treatment of Somnolence in Parkinson's Disease: Results of n-of-1 Trials. Frontiers in Neurology. 2016 Mar 8;7:27.

Frost KL, Carey JR, Broback TW, Carlson NL, Daggett CA, Dalbec MM, Mueller BA. N-of-1 trial in person with pontine stroke receiving repetitive transcranial magnetic stimulation to improve hand function. Journal of neuroimaging in psychiatry & neurology. 2017;2(2):36.

Germini F, Coerezza A, Andreinetti L, Nobili A, Rossi PD, Mari D, Guyatt G, Marcucci M. N-of-1 randomized trials of ultra-micronized palmitoylethanolamide in older patients with chronic pain. Drugs & Aging. 2017 Dec;34(12):941-52.

Gurfinkel R, Leushner J, Paul TL, Mahon JL. Use of an N-of-1 Trial to Assess Levothyroxine Intolerance in a Patient with Hypothyroidism. Canadian Journal of Diabetes. 2013 Oct 1;37:S24-5.

Gursli S, Sandvik L, Bakkeheim E, Skrede B, Stuge B. Evaluation of a novel technique in airway clearance therapy–Specific Cough Technique (SCT) in cystic fibrosis: A pilot study of a series of N-of-1 randomised controlled trials. SAGE Open Medicine. 2017 Mar 15;5:2050312117697505.

Howard JP, Wood F, Finegold J, Nowbar A, Thompson D, Arnold AD, Rajkumar C, Christopher SB, Connolly S, Cegla J, Sever PS. A Three-arm N-of-1 Trial With Statin, Placebo and Tablet Free Periods, to Verify Side Effects and Identify Their Cause: The SAMSON Trial. InCIRCULATION 2020 Dec 15 (Vol.

142, No. 24, pp. E480-E481). TWO COMMERCE SQ, 2001 MARKET ST, PHILADELPHIA, PA 19103 USA: LIPPINCOTT WILLIAMS & WILKINS.

Hsia CC, Mahon JL, Seitelbach M, Chia J, Zou G, Chin-Yee IH. Use of n-of-1 (single patient) trials to assess the effect of age of transfused blood on health-related quality of life in transfusion-dependent patients. Transfusion. 2016 May;56(5):1192-200.

Huang H, Yang P, Wang J, Wu Y, Zi S, Tang J, Wang Z, Ma Y, Zhang Y. Investigation into the individualized treatment of traditional Chinese medicine through a series of N-of-1 trials. Evidence-Based Complementary and Alternative Medicine. 2018 Jan 1;2018.

Huang H, Zhang Y, Yang P, Xue J, Tang J, Guyatt G. The Pilot Study of Traditional Chinese Medicine in the Treatment of Stable Bronchiectasis by N-of-1 Trials. The Journal of Alternative and Complementary Medicine. 2014 May 1;20(5):A39-.

Hyam JA, Pereira EA, McCulloch P, Javed S, Plaha P, Mooney L, Forrow BA, Joint C, Whone A, Gill S, Glasziou P. Implementing novel trial methods to evaluate surgery for essential tremor. British Journal of Neurosurgery. 2015 May 4;29(3):334-9.

Janssen DA, Hoogenboom T, Heesakkers JP. Novel and feasible methodology to obtain LE1 evidence for BPS treatment. Implementing the combined N-of-1 trial design. European Urology Supplements. 2019 Mar 1;18(1):e1828.

Jensen EA, Zhang H, Feng R, Dysart K, Nilan K, Munson DA, Kirpalani H. Individualising care in severe bronchopulmonary dysplasia: a series of N-of-1 trials comparing transpyloric and gastric feeding. Archives of Disease in Childhood-Fetal and Neonatal Edition. 2020 Jul 1;105(4):399-404.

Joy TR, Monjed A, Zou GY, Hegele RA, McDonald CG, Mahon JL. N-of-1 (single-patient) trials for statin-related myalgia. Annals of internal medicine. 2014 Mar 4;160(5):301-10.

Kaplan HC, Opipari-Arrigan L, Schmid CH, Schuler CL, Saeed S, Braly KL, Burgis JC, Nguyen K, Pilley S, Stone J, Woodward G. Evaluating the comparative effectiveness of two diets in pediatric inflammatory bowel disease: a study protocol for a series of N-of-1 trials. InHealthcare 2019 Nov 1 (Vol. 7, No. 4, p. 129). MDPI.

Kronish IM, Cheung YK, Julian J, Parsons F, Lee J, Yoon S, Valdimarsdottir H, Green P, Suls J, Hershman DL, Davidson KW. Clinical usefulness of bright white light therapy for depressive symptoms in cancer survivors: Results from a series of personalized (n-of-1) trials. InHealthcare 2019 Dec 30 (Vol. 8, No. 1, p. 10). MDPI.

Lampert A, Tesarz J, Volkert AK, Wesche D, Wild B, Schwab M, Haefeli WE. An N-of-1 Trial as an Individualized Withdrawal Treatment Approach to Psychological Methylphenidate Dependence. Psychotherapy and psychosomatics. 2014 Jan 1;83(6):379-81.

Lampert A, Tesarz J, Volkert AK, Wesche D, Wild B, Schwab M, Haefeli WE. An N-of-1 Trial as an Individualized Withdrawal Treatment Approach to Psychological Methylphenidate Dependence. Psychotherapy and psychosomatics. 2014 Jan 1;83(6):379-81.

Lee S, Lim N, Choi SM, Kim S. Validation study of Kim's Sham needle by measuring facial temperature: an N-of-1 randomized double-blind placebo-controlled clinical trial. Evidence-Based Complementary and Alternative Medicine. 2012 Jan 1;2012.

Lipka AF, Vrinten C, van Zwet EW, Schimmel KJ, Cornel MC, Kuijpers MR, Hekster YA, Weinreich SS, Verschuuren JJ. Ephedrine treatment for autoimmune myasthenia gravis. Neuromuscular Disorders. 2017 Mar 1;27(3):259-65.

Marcucci M, Germini F, Coerezza A, Andreinetti L, Bellintani L, Nobili A, Rossi PD, Mari D. Efficacy of ultra-micronized palmitoylethanolamide (um-PEA) in geriatric patients with chronic pain: study protocol for a series of N-of-1 randomized trials. Trials. 2016 Dec;17(1):1-2.

McGarry ME, Illek B, Ly NP, Zlock L, Olshansky S, Moreno C, Finkbeiner WE, Nielson DW. In vivo and in vitro ivacaftor response in cystic fibrosis patients with residual CFTR function: N-of-1 studies. Pediatric pulmonology. 2017 Apr;52(4):472-9.

Nikles J, O'Sullivan JD, Mitchell GK, Smith SS, McGree JM, Senior H, Dissanyaka N, Ritchie A. Protocol: Using N-of-1 tests to identify responders to melatonin for sleep disturbance in Parkinson's disease. Contemporary Clinical Trials Communications. 2019 Sep 1;15:100397.

Nikles J, Mitchell GK, Hardy J, Agar M, Senior H, Carmont SA, Schluter PJ, Good P, Vora R, Currow D. Testing pilocarpine drops for dry mouth in advanced cancer using n-of-1 trials: A feasibility study. Palliative Medicine. 2015 Dec;29(10):967-74.

Ong KS, Carlin JB, Fahey M, Freeman JL, Scheffer IE, Gillam L, Anderson M, Huque MH, Legge D, Dirnbauer N, Lilley B. Protocol for a single patient therapy plan: A randomised, double-blind, placebo-controlled N-of-1 trial to assess the efficacy of cannabidiol in patients with intractable epilepsy. Journal of paediatrics and child health. 2020 Dec;56(12):1918-23.

Persson MS, Stocks J, Sarmanova A, Fernandes G, Walsh DA, Doherty M, Zhang W. Individual responses to topical ibuprofen gel or capsaicin cream for painful knee osteoarthritis: a series of n-of-1 trials. Rheumatology. 2021 May;60(5):2231-7.

Riggare S, Unruh KT, Sturr J, Domingos J, Stamford JA, Svenningsson P, Hägglund M. Patient-driven N-of-1 in Parkinson's disease. Methods of information in medicine. 2017 Jan;56(S 01):e123-8.

Roustit M, Giai J, Gaget O, Khouri C, Mouhib M, Lotito A, Blaise S, Seinturier C, Subtil F, Paris A, Cracowski C. On-demand sildenafil as a treatment for Raynaud phenomenon: a series of N-of-1 trials. Annals of Internal Medicine. 2018 Nov 20;169(10):694-703.

Samuel JP, Samuels JA, Brooks LE, Bell CS, Pedroza C, Molony DA, Tyson JE. Comparative effectiveness of antihypertensive treatment for older children with primary hypertension: study protocol for a series of n-of-1 randomized trials. Trials. 2016 Dec;17(1):1-9.

Santos C, Weaver DF. Topically applied linoleic/linolenic acid for chronic migraine. Journal of Clinical Neuroscience. 2018 Dec 1;58:200-1.

Seitelbach M, Chin-Yee IH, Kinney J, Chia J, Ormond K, Mahon J, Zou G, Stitt L, Hsia CC. Age of Blood Does Not Affect Quality of Life and Hemoglobin: N–of–1 Trials in Transfusion-Dependent Patients. Blood. 2011 Nov 18;118(21):3369.

Senior HE, McKinlay L, Nikles J, Schluter PJ, Carmont SA, Waugh MC, Epps A, Lloyd O, Mitchell GK. Central nervous system stimulants for secondary attention deficit-hyperactivity disorder after paediatric traumatic brain injury: a rationale and protocol for single patient (n-of-1) multiple cross-over trials. BMC pediatrics. 2013 Dec;13(1):1-7.

Senior HE, Mitchell GK, Nikles J, Carmont SA, Schluter PJ, Currow DC, Vora R, Yelland MJ, Agar M, Good PD, Hardy JR. Using aggregated single patient (N-of-1) trials to determine the effectiveness of

psychostimulants to reduce fatigue in advanced cancer patients: a rationale and protocol. BMC Palliative Care. 2013 Dec;12(1):1-6.

Sierra-Arango F, Castaño DM, Forero JD, Pérez-Riveros ED, Ardila Duarte G, Botero ML, Cárdenas A, la Hoz-Valle D. A randomized placebo-controlled N-of-1 trial: the effect of proton pump inhibitor in the management of gastroesophageal reflux disease. Canadian Journal of Gastroenterology and Hepatology. 2019 Dec 12;2019.

Smith JH, Potter JL, Robblee JV. What Is the Optimal Placebo for Clinical Trials of Nerve Blocks in Headache?: An N-of-1 Study of Lactated Ringer's. Headache: The Journal of Head and Face Pain. 2019 Jan;59(1):108-9.

Sniehotta FF, Presseau J, Hobbs N, Araújo-Soares V. Testing self-regulation interventions to increase walking using factorial randomized N-of-1 trials. Health Psychology. 2012 Nov;31(6):733.

Stunnenberg BC, Raaphorst J, Groenewoud HM, Statland JM, Griggs RC, Woertman W, Stegeman DF, Timmermans J, Trivedi J, Matthews E, Saris CG. Effect of mexiletine on muscle stiffness in patients with nondystrophic myotonia evaluated using aggregated N-of-1 trials. Jama. 2018 Dec 11;320(22):2344-53.

Sun M, Chai L, Lu F, Zhao Y, Li Q, Cui B, Gao R, Liu Y. Efficacy and safety of Ginkgo Biloba pills for coronary heart disease with impaired glucose regulation: study protocol for a series of N-of-1 randomized, double-blind, placebo-controlled trials. Evidence-Based Complementary and Alternative Medicine. 2018 Jan 1;2018.

Taye I, Bradbury J, Grace S, Avila C. Probiotics for pain of osteoarthritis; An N-of-1 trial of individual effects. Complementary therapies in medicine. 2020 Nov 1;54:102548.

Tian Y, Ma Y, Fu Y, Zheng JS. Application of n-of-1 clinical trials in personalized nutrition research: a trial protocol for Westlake N-of-1 Trials for Macronutrient Intake (WE-MACNUTR). Current Developments in Nutrition. 2020 Sep;4(9):nzaa143.

Tison F, Nègre-Pagès L, Meissner WG, Dupouy S, Li Q, Thiolat ML, Thiollier T, Galitzky M, Ory-Magne F, Milhet A, Marquine L. Simvastatin decreases levodopa-induced dyskinesia in monkeys, but not in a randomized, placebo-controlled, multiple cross-over ("n-of-1") exploratory trial of simvastatin against levodopa-induced dyskinesia in Parkinson's disease patients. Parkinsonism & related disorders. 2013 Apr 1;19(4):416-21.

Tsiormpatzis S. Effects of shiatsu on the health-related quality of life of a person with secondary progressive multiple sclerosis: A mixed methods N-of-1 trial within a whole systems research case study. European Journal of Integrative Medicine. 2019 Dec 1;32:101006.

Vrinten C, Lipka AF, van Zwet EW, Schimmel KJ, Cornel MC, Kuijpers MR, Hekster YA, Weinreich SS, Verschuuren JJ. Ephedrine as add-on therapy for patients with myasthenia gravis: protocol for a series of randomised, placebo-controlled n-of-1 trials. BMJ open. 2015 Jul 1;5(7):e007863.

Weijma R, Merkus E, van Vlijmen B, Groenewoud H, van der Wilt GJ, van Engelen B, Raaphorst J, Saris C, Stunnenberg B. Individual N-of-1 Trial Salbutamol Versus Placebo as Attack Treatment in a Patient with Hyperkalemic Periodic Paralysis (P3. 455).

Weng S, Fan Z, Qiu G, Liu F, Huang L, Li J, Jiang X, Song Z, Gao Y, Zhong Z, He L. Therapeutic efficacy and immunoregulatory effect of Qiangji Jianli Capsule for patients with myasthenia gravis: Study protocol for a series of randomized, controlled N-of-1 trials. Medicine. 2020 Dec 12;99(51).

Wood FA, Howard JP, Finegold JA, Nowbar AN, Thompson DM, Arnold AD, Rajkumar CA, Connolly S, Cegla J, Stride C, Sever P. N-of-1 trial of a statin, placebo, or no treatment to assess side effects. New England Journal of Medicine. 2020 Nov 26;383(22):2182-4.

Yuhong H, Qian L, Yu L, Yingqiang Z, Yanfen L, Shujing Y, Shufang Q, Lanjun S, Shuxuan Z, Baohe W. An n-of-1 trial service in clinical practice: testing the effectiveness of Liuwei Dihuang decoction for Kidney-Yin deficiency syndrome. Evidence-Based Complementary and Alternative Medicine. 2013 Jan 1;2013.

Samuel JP, Tyson JE, Green C, Bell CS, Pedroza C, Molony D, Samuels J. Treating hypertension in children with n-of-1 trials. Pediatrics. 2019 Apr 1;143(4).