# DELTA2 guidance on choosing the target difference and undertaking and reporting the sample size calculation for a randomised controlled trial

***Submitted for consideration as a Research Methods and Reporting article***

***Version 10: 26-7-18***

****

# Jonathan A Cook, Associate Professor, jonathan.cook@ndorms.ox.ac.uk1, Steven A Julious, s.a.julious@sheffield.ac.uk, Professor2, William Sones william.sones@ndorms.ox.ac.uk, Statistician1, Lisa V Hampson, lisa.hampson@novartis, Associate Director3,4, Catherine Hewitt, catherine.hewitt@york.ac.uk, Professor5, Jesse A Berlin, jberlin@its.jnj.com, Vice President and Global Head of Epidemiology6, Deborah Ashby, deborah.ashby@imperial.ac.uk, Co-Director7,  Richard Emsley, Richard.Emsley@kcl.ac.uk, Professor8, Fergusson Dean A, dafergusson@ohri.ca, Senior Scientist & Director9, Stephen J Walters, s.j.walters@sheffield.ac.uk, Professor2, Edward CF Wilson, ed.wilson@medschl.cam.ac.uk, Senior Research Associate in Health Economics10, Graeme MacLennan, g.maclennan@abdn.ac.uk, Director & Professor11, Nigel Stallard, N.Stallard@warwick.ac.uk, Professor12, Joanne C Rothwell, j.c.rothwell@sheffield.ac.uk, PhD student2, Martin Bland, martin.bland@york.ac.uk, Professor4, Louise Brown, l.brown@ucl.ac.uk, Senior Statistician13, Craig R Ramsay, c.r.ramsay@abdn.ac.uk, Director & Professor14, Andrew Cook, andrewc@soton.ac.uk, Consultant in Public Health Medicine and Fellow in Health Technology Assessment15, David Armstrong, david.armstrong@kcl.ac.uk, Professor16, Doug Altman, doug.altman@csm.ox.ac.uk, Professor 1, Luke D Vale, luke.vale@newcastle.ac.uk, Professor17.

**ADDRESSES**

1 Centre for Statistics in Medicine,Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Botnar Research Centre, Nuffield Orthopaedic Centre, Windmill Rd, Oxford, OX3 7LD

2 Medical Statistics Group, ScHARR
The University of Sheffield
Regent Court, 30 Regent Street
SHEFFIELD S1 4DA

3 Statistical Methodology & Consulting, Novartis, Basel, Switzerland

# 4 Department of Mathematics and Statistics

Lancaster University

Lancaster, LA1 4YF

# 5 Department of Health Sciences, Seebohm Rowntree Building

University of York, Heslington, York, YO10 5DD, UK

# 6 [Johnson & Johnson](http://www.zoominfo.com/c/Johnson-%26-Johnson/61212649)

# 1125 Trenton-Harbourton Road

# Titusville, New Jersey 08933

# United States

7 Imperial Clinical Trials Unit, Deputy Head, School of Public Health,

Imperial College London

Stadium House, 68 Wood Lane

London W12 7RH

8 Department of Biostatistics and Health InformaticsInstitute of Psychiatry, Psychology and Neuroscience

King’s College London

De Crespigny Park

Denmark Hill

London, SE5 8AF

# 9 Clinical Epidemiology Program,

# Ottawa Hospital Research Institute,

# 501 Smyth Box 511,

# Ottawa, ON, K1H 8L6,

# Canada

10 Cambridge Centre for Health Services Research & Cambridge Clinical Trials Unit
University of Cambridge
Institute of Public Health
Forvie Site, Robinson Way
Cambridge, CB2 0SR

11 The Centre for Healthcare Randomised Trials (CHaRT)

Health Sciences Building

University of Aberdeen

Foresterhill

Aberdeen, AB25 2ZD

12 Warwick Medical School - Statistics and Epidemiology
University of Warwick
Coventry, CV4 7AL

13 MRC Clinical Trials Unit at UCL

Institute of Clinical Trials & Methodology

2nd Floor, 90 High Holborn

London, WC1V 6LJ

14 Health Services Research Unit

University of Aberdeen

Health Sciences Building

Foresterhill

Aberdeen, AB25 2ZD

15 Wessex Institute, University of Southampton

Alpha House, Enterprise Road,

Southampton, SO16 7NS

16 School of Population Health & Environmental Sciences

Faculty of Life Sciences and Medicine
King's College London
Addison House, Guy’s Campus,

London, SE1 1UL

17 Health Economics Group

Institute of Health & Society

Newcastle University

Newcastle upon Tyne

UK, NE2 4AX

\* **Correspondance to: J Cook jonathan.cook@ndorms.ox.ac.uk**

**MANUSCRIPT WORD COUNT 2449**

**Keywords**

Target difference, clinically important difference, sample size, guidance, randomised trial, effect size, realistic difference

|  |
| --- |
| ***Standfirst*** The randomised controlled trial (RCT) is considered the gold standard to assess comparative clinical efficacy and effectiveness and can be a key source of data for estimating cost-effectiveness. Central to the design of a RCT is an *a-priori* sample size calculation which ensures the study has a high probability of achieving its pre-specified main objective. Beyond pure statistical or scientific concerns, it is ethically imperative that an appropriate number of study participants be recruited, to avoid imposing the burdens of a clinical trial on more patients than necessary. The scientific concern is satisfied and the ethical imperative further addressed by specifying a target difference between treatments that is considered realistic and/or important by one or more key stakeholder groups. The sample size calculation ensures that the trial will have the required statistical power to identify whether a difference of a particular magnitude exists. The key messages from the DELTA2 guidance on determining the target difference and sample size calculation for a RCT are presented. Recommendations for the subsequent reporting of the sample size calculation are also provided. |

|  |
| --- |
| Summary points * Central to the design of a RCT is an *a-priori* sample size calculation, which ensures there is a high probability of the study achieving its pre-specified main objective.
* Getting the sample size wrong can lead to a study which is unable to inform clinical practice (hence directly or indirectly harming patients), or could expose excess patients to the uncertainty inherent in a clinical trial.
* The target difference between treatments which is considered realistic and/or important by one or more key stakeholder groups plays a critical part in the sample size calculation of a randomised controlled trial.
* Guidance on how to choose the target difference and to undertake a sample size calculation for funders and researchers is presented.
* 10 recommendations are made regarding choosing the target difference and undertaking a sample size calculation along with recommended reporting items for trial proposal, protocols and results papers.
 |

**Introduction**

Properly conducted, the randomised controlled trial (RCT) is considered to be the gold standard for assessing the comparative clinical efficacy and effectiveness of healthcare interventions, as well as providing a key source of data for estimating cost-effectiveness.1 RCTs are routinely used to evaluate a wide range of treatments and have been successfully used in a variety of health and social care settings. Central to the design of a RCT is an *a-priori* sample size calculation, which ensures the study has a high probability of achieving its pre-specified objective.

The difference between groups used to calculate a sample size for the trial, the “target difference”, is the magnitude of difference in the outcome of interest that the RCT is designed to reliably detect. Reassurance in this regard is typically confirmed by having a sample size which has a sufficiently high level of statistical power (typically 80 or 90%) for detecting a difference as big as the target difference, while setting the statistical significance at the level planned for the statistical analysis (usually this is the 2-sided 5% level). A comprehensive methodological review conducted by the original DELTA (Difference ELicitation in TriAls) group2 3 highlighted the available methods and limitations in current practice. It showed that despite there being many different approaches available, some are used only rarely in practice.4 The initial DELTA guidance does not fully meet the needs of funders and researchers. The overall aim of the DELTA2 project, commissioned by the UK Medical Research Council (MRC)/National Institute for Health Research (NIHR) Methodology Research Programme, and described here, was to produce updated guidance for researchers and funders on specifying and reporting the target difference (“effect size”) in the sample size calculation of a RCT. The process of developing the new guidance is summarised in the following section. We then summarise the relevant considerations, key messages and recommendations for determining and reporting a RCT’s sample size calculation (Boxes 1 and 2).

**Development of the guidance**

## The DELTA2 guidance is the culmination of a five stage process to meet the stated project objectives (see Figure 1) which included two literature reviews of existing funder guidance and recent methodological literature, a Delphi process to engage with a wider group of stakeholders, a 2 day workshop and finalising the core guidance.

**Figure 1 DELTA2 project components of work**

The literature review was conducted between April and December 2016 (searching up to Apr 2016). The Delphi study had two rounds, one held in 2016 before a two-day workshop held in Oxford (September 2016) and one following it between August and November 2017. The general structure of the guidance was devised at the workshop. It was substantially revised based upon feedback from stakeholders received through the Delphi process. In addition, stakeholder engagement events were held at various meetings throughout the development of the guidance (the Society for Clinical Trials (SCT) meeting, and Statisticians in the Pharmaceutical Industry (PSI) conferences both in May 2017, Joint Statistical Meeting (JSM) in August 2017 and a Royal Statistical Society (RSS) Reading local group meeting in September 2017). These interactive sessions provided feedback on the scope (in 2016) and then draft guidance (in 2017). The core guidance was provisionally finalised in October 2017 and reviewed by the funders’ representatives for comment (MRP advisory group). The guidance was further revised and finalised in February 2018. The full guidance document incorporating case studies and relevant appendices is available here.5 Further details on the findings of the Delphi study and the wider engagement with stakeholders are reported elsewhere.6 The guidance and key messages are summarised in the remainder of this paper.

**The target difference and RCT sample size calculations**

The role of the sample size calculation is to determine the number of patients required in order that the planned analysis of the primary outcome is likely to be informative. It is typically achieved by specifying a target difference for the key (primary) outcome which can be reliably detected and the required sample size calculated. Within this summary paper we restrict considerations to the most common trial design addressing a superiority question (one which assumes no difference between treatments and looks for a difference) though the full guidance considers equivalence and non-inferiority designs which invert the hypothesis and how the use of the target difference differs for such designs.5

The precise research question that the trial is primarily set up to answer will determine what needs to be estimated in the planned primary analysis, this is known formally as the ‘estimand’. A key part of defining this is choosing the primary outcome, which requires careful consideration. The target difference should be a difference that is appropriate for that estimand.7-10 This is typically (for superiority trials) an “intention to treat” or treatment policy estimand i.e. according to the randomised groups irrespective of subsequent compliance with the treatment allocation. Other analyses that address different estimands8 9 11 of interest (e.g. those based on the effect upon receipt of treatment and the absence of non-compliance) could also inform the choice of sample size. Different stakeholders can have somewhat differing perspectives on the appropriate target difference12. However, a key principle is that the target difference should be one that would be viewed as important by at least one (preferably more) key stakeholder groups i.e. patients, health professionals, regulatory agencies, and healthcare funders. In practice, the target difference is not always formally considered and in many cases appears, at least from trial reports, to be determined upon convenience, the research budget, and/or some other informal basis.13 The target difference can be expressed as an absolute difference (e.g., mean difference or difference in proportions) or a relative difference (e.g., hazard or risk ratio), and it is also often referred to, rather imprecisely, as the trial “effect size”.

Statistical sample size calculation is far from an exact science.14 First, investigators typically make assumptions that are a simplification of the anticipated analysis. For example, the impact of adjusting for baseline factors is very difficult to quantify upfront, and even though the analysis is intended to be an adjusted one (for example, when randomisation has been stratified or minimised),15 the sample size calculation is often conducted based on an unadjusted analysis. Second, the calculated sample size can be very sensitive to the assumptions made in the calculations such that a small change in one of the assumptions can lead to substantial change in the calculated sample size. Often a simple formula can be used to calculate the required sample size. The formula varies according to the type of outcome, how the target difference is expressed (e.g. a risk ratio versus a difference in proportions),and somewhat implicitly, the design of the trial and the planned analysis. Typically, a sample size formula can be used to calculate the required number of observations in the analysis set, which varies depending on the outcome and the intended analysis. In some situations, ensuring the sample size is sufficient for more than one planned analysis may be appropriate.

When deciding upon the sample size for a RCT, it is necessary to balance the risk of incorrectly concluding there is a difference when no actual difference between the treatments exists, with the risk of failing to identify a meaningful treatment difference when the treatments do differ. Under the conventional approach, referred to as the statistical hypothesis testing framework16, the probabilities of these two errors are controlled by setting the significance level (Type I error) and statistical power (1 minus Type II error) at appropriate levels (typical values are 2 sided 5% significance and 80 or 90% power respectively). Once these two inputs have been set, the sample size can be determined given the magnitude of the between group difference in the outcome it is desired to detect (the target difference). The calculation (reflecting the intended analysis) is conventionally done on the basis of testing for a difference of any magnitude. As a consequence, it is essential when interpreting the analysis of a trial to consider the uncertainty in the estimate, which is reflected in the confidence interval. A key question of interest is what magnitude of difference can be ruled out. The expected (predicted) width of the confidence interval can be determined for a given target difference and sample size calculation which is a helpful further aid in making an informed choice about this part of a trial’s design.17 Other statistical and economic approaches to calculating the sample size have been proposed such as precision and Bayesian based approaches,16 18-21 and value of information analysis,22 though they are not at present commonly applied.18

The required sample size is very sensitive to the target difference. Under the conventional approach, halving the target difference quadruples the sample size for a two arm 1:1 parallel group superiority trial with a continuous outcome.23 Appropriate sample size formulae vary depending upon the proposed trial design and statistical analysis, although the overall approach is consistent. In more complex scenarios, simulations may be used but the same general principles hold. It is prudent to undertake sensitivity calculations to assess the potential impact of misspecification of key assumptions (such as the control response rate for a binary outcome or the anticipated variance of a continuous outcome).

The sample size calculation and the target difference, if well specified, help provide reassurance that the trial is likely to detect a difference at least as large as the target difference in terms of comparing the primary outcome between treatments. Failure to clarify sufficiently what is important and realistic at the design stage can lead to subsequent sample size revisions, an unnecessarily inconclusive trial due to lack of statistical precision, or to ambiguous interpretation of the findings.24 25 When specifying the target difference with a definitive trial in mind, the following guidance should be considered.

**Specifying the target difference for a randomised controlled trial**

Different statistical approaches can be taken to specify the target difference and calculate the sample size but the general principles are the same. To aid those new to the topic and to encourage better practice and reporting regarding the specification of the target difference for a RCT, a series of *recommendations* is provided in Boxes 1 and 2. Seven broad types of methods can be used to justify the choice of a particular value as the target difference: these are summarised in Box 3.

Broadly speaking, two different approaches can be taken to specify the target difference for a RCT. A difference that is considered to be:

* *important* to one or more stakeholder groups
* *realistic* (plausible),based on either existing evidence, and/or expert opinion.

A very large literature exists on defining and justifying a (clinically) important difference, particularly for quality of life outcomes.26-28 In a similar manner, discussions of the relevance of estimates from existing studies are also common; there are a number of potential pitfalls to their use, which requires careful consideration of how they should inform the choice of the target difference.2 It should be noted that it has been argued that a target difference should always be both important and realistic.29 This would seem particularly apt when designing a definitive (Phase III) superiority RCT. In a RCT sample size calculation, the target difference between the treatment groups, strictly relates to a group level difference for the anticipated study population. However, the difference in an outcome that is important to an individual might differ from the corresponding value at the population level. More extensive consideration of the variations in approach is provided elsewhere.3 30

**Reporting the sample size calculation**

The approach taken when determining the sample size and the assumptions made should be clearly specified. This includes all the inputs and formula or simulation results, so that it is clear what the sample size was based upon. This information is critical for reporting transparency, allows the sample size calculation to be replicated, and clarifies the primary (statistical) aim of the study. Under the conventional approach with a standard (1:1 allocation two arm parallel group superiority) trial design and unadjusted statistical analysis, the core items needed to be stated are the primary outcome, the target difference appropriately specified according to the outcome type, the associated “nuisance” parameter (parameters that together with the target difference uniquely specifies the difference on the original outcome scale, e.g., for a binary primary outcome this is the event rate in the control group) and the statistical significance and power. More complicated designs can have additional inputs that also need considered, like the intra-cluster correlation for a cluster randomised design.

A set of core items should be reported in all key trial documents (grant applications, protocols and main results papers) to ensure reproducibility and plausibility of the sample size calculation. The full list of recommended core items are given in Box 2 which is an update of the previously-proposed list.31 When the sample size calculation deviates from the conventional approach, whether by research question or statistical framework, the core reporting set may be modified to provide sufficient detail to ensure the sample size calculation is reproducible and the rationale for choosing the target difference is transparent. However, the key principles remain the same. Where the sample size is determined based upon a series of simulations, this would need to be described in sufficient detail to enable equivalent level of transparency and assessment. Additional items to give more explanation of the rationale should be provided where space allows (e.g. grant applications and trial protocols). Trial result publications can then reference these documents if sufficient space is not available to provide a full description.

**Discussion**

Researchers are faced with a number of difficult decisions when designing a RCT, the most important of which are the choice of trial design, primary outcome and sample size. The latter is largely driven by the choice of the target difference, although other aspects of sample size determination also contribute.

The DELTA2 guidance on specifying a target difference and undertaking and reporting the sample size calculation for a RCT was developed in response to a growing recognition from funders, researchers, as well as other key stakeholders (such as patients and the respective clinical communities), that there is a real need for practical and accessible advice to inform a difficult decision. The new guidance document therefore aims to bridge the gap between the existing (limited) guidance and this growing need.

The key message for researchers is the need to be more explicit about the rationale and justification of the target difference when undertaking and reporting a sample size calculation.. Increasing focus is being placed upon the target difference in the clinical interpretation of the trial result, whether statistically significant or not. There is a need to improve on the specification of the area target difference and the reporting of this.

**Acknowledgements**

This project was funded by the MRC-NIHR Methodology Research Programme in the UK in response to a commissioned call to lead a workshop on this topic in order to produce guidance. The members of the original DELTA (Difference ELicitation in TriAls)2 group were:

Professor Jonathan Cook, Professor Doug Altman, Dr Jesse Berlin, Professor Martin Bland, Professor Richard Emsley, Dr Dean Fergusson, Dr Lisa Hampson, Professor Catherine Hewitt, Professor Craig Ramsay, Miss Joanne Rothwell, Dr Robert Smith, Dr William Sones, Professor Luke Vale, Professor Stephen Walters, and Professor Steve Julious.

As part of the process of development of the guidance a two day workshop was held in Oxford in September 2016. The workshop participants were:

Professor Doug Altman, Professor David Armstrong, Professor Deborah Ashby, Professor Martin Bland, Dr Andrew Cook, Professor Jonathan Cook, Dr David Crosby, Professor Richard Emsley, Dr Dean Fergusson, Dr Andrew Grieve, Dr Lisa Hampson, Professor Catherine Hewitt, Professor Steve Julious, Professor Graeme MacLennan, Professor Tim Maughan, Professor Jon Nicholl, Dr José Pinheiro, Professor Craig Ramsay, Miss Joanne Rothwell, Dr William Sones, Professor Nigel Stallard, Professor Luke Vale, Professor Stephen Walters, and Dr Ed Wilson.

The authors would like to acknowledge and thank the participants in the Delphi study and the one-off engagement sessions with various groups including the Society for Clinical Trials, PSI and Joint Statistical Meeting conference session attendees, along with the other workshop participants who kindly provided helpful input and comments on the scope and content of this document. We would also like to thank in particular Dr Robert Smith in his role as a member of the public who provided helpful public perspective and input in the workshop and also development and revision of the guidance document. The authors would like to thank Stefano Vezzoli for in-depth comments which helped to refine this document. Finally, the authors acknowledge the helpful feedback from the MRC MRP advisory panel which commissioned the study and provided feedback on a draft version.

**DECLARATIONS**

**Author contributions**

JAC and SAJ conceived of the idea and drafted the initial version of the manuscript. WS, LVH, CH, JAB, DA, RE, DAF, SJW, ECHW, GM, NS, JCR, MB, LB, CRR, AC, DA, DA and LDV contributed to the development of the guidance and commented on the draft manuscript. All authors have read and approved the final version. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted

**Competing interests**

All authors have completed the Unified Competing Interest form at [www.icmje.org/coi\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) (available on request from the corresponding author). JAB is employee of J&J and holds shares in this company. LVH is an employee of Novartis. The authors declare grant funding from the MRC and NIHR UK for this work. All of the other authors have been involved in design and conducting randomised trials through their roles. The authors declare that they have no other financial relationships that might have an interest in the submitted work; and all authors declare they have no other non-financial interests that may be relevant to the submitted work.

**Ethics approval and consent to participate**

Ethics approval for the Delphi study which is part of the DELTA2 project was sought and received from the University of Oxford’s Medical Sciences Inter-divisional Research Ethics Committee (IDREC - R46815/RE001). Informed consent was obtained for all participants in the Delphi study.

**Funding**

Funding for this work was received from the MRC-NIHR UK Methodology Research Panel in response to an open commissioned call for an Effect Size Methodology State-of-the-art Workshop. The Health Services Research Unit, Institute of Applied Health Sciences (University of Aberdeen), is core-funded by the Chief Scientist Office of the Scottish Government Health and Social Care Directorates. The funders had no involvement in study design, collection, analysis and interpretation of data, reporting or the decision to publish.

**Copyright**

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, [*a worldwide licence*](http://www.bmj.com/sites/default/files/BMJ%20Author%20Licence%20March%202013.doc) to the Publishers and its licensees in perpetuity, in all forms, formats and media (whether known now or created in the future), to i) publish, reproduce, distribute, display and store the Contribution, ii) translate the Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) licence any third party to do any or all of the above.

**Paper's provenance**

This paper summarised the key findings of the new guidance produced by the DELTA2 study. MRC-NIHR UK Methodology Research Panel in response to an open commissioned call for an Effect Size Methodology State-of-the-art Workshop. The authors are all researchers who have been involved in randomised trials of varying types with most involved for 10 plus years. They have varying backgrounds and have worked in a range of clinical areas and on both academic and industry funded studies.

|  |
| --- |
| **Box 1 DELTA2 recommendations for for undertaking a sample size calculation for a RCT**1. Begin by searching for relevant literature to inform the specification of the target difference. Relevant literature can:
	1. relate to a candidate primary outcome and/or the comparison of interest, and;
	2. inform what is an important and/or realistic difference for that outcome, comparison and population.
2. Candidate primary outcomes should be considered in turn, and the corresponding sample size explored. Where multiple candidate outcomes are considered, the choice of the primary outcome and target difference should be based upon consideration of the views of relevant stakeholders groups (for example, patients), as well as the practicality of undertaking such a study with the required sample size. The choice should not be based solely on which yields the minimum sample size. Ideally, the final sample size will be sufficient for all key outcomes though this is not always practical.
3. The importance of observing a particular magnitude of a difference in an outcome, with the exception of mortality and other serious adverse events, cannot be presumed to be self-evident. Therefore, the target difference for all other outcomes requires additional justification to infer importance to a stakeholder group.
4. The target difference for a definitive (e.g. Phase III) trial should be one considered to be important to at least one key stakeholder group.
5. The target difference does not necessarily have to be the minimum value that would be considered important if a larger difference is considered a realistic possibility or would be necessary to alter practice.
6. Where additional research is needed to inform what would be an important difference, the anchor and opinion seeking methods are to be favoured. The distribution method should not be used. Specifying the target difference based solely upon a Standardised Effect Size approach should be considered a last resort though it may be helpful as a secondary approach.
7. Where additional research is needed to inform what would be a realistic difference, the Opinion Seeking and the Review of the Evidence Base methods are recommended. Pilot trials are typically too small to inform what would be a realistic difference and primarily address other aspects of trial design and conduct.
8. Use existing studies to inform the value of key “nuisance” parameters which are part of the sample size calculation. For example, a pilot trial can be used to inform the choice of the standard deviation value for a continuous outcome and the control group proportion for a binary outcome, along with other relevant inputs such as the amount of missing outcome data.
9. Sensitivity analyses, which consider the impact of uncertainty around key inputs (e.g. the target difference and the control group proportion for a binary outcome) used in the sample size calculation, should be carried out.
10. Specification of the sample size calculation, including the target difference, should be reported according to the guidance for reporting items (see below) when preparing key trial documents (grant applications, protocols and result manuscripts).
 |

|  |
| --- |
| **Box 2 DELTA2 recommended reporting items for the sample size calculation of a RCT with a superiority question** |
| ***Core items***1. Primary outcome (and any other outcome on which the calculation is based)
	1. If a primary outcome is not used as the basis for the sample size calculation, state why.
2. Statistical significance level and power
3. Express the target difference according to outcome type
	1. Binary – state the target difference as an absolute and/or relative effect, along with the intervention and control group proportions. If both an absolute and a relative difference are provided, clarify if either takes primacy in terms of the sample size calculation.
	2. Continuous – state the target mean difference on the natural scale, the common SD and the standardised effect size (mean difference divided by the SD).
	3. Time-to-event – state the target difference as an absolute and/or relative difference; provide the control group event proportion; the planned length of follow-up; and the intervention and control group survival distributions and the accrual time (if assumptions regarding them are made). If both an absolute and relative difference are provided for a particular time point, clarify if either takes primacy in terms of the sample size calculation.
4. Allocation ratio
	1. If an unequal ratio is used, the reason for this should be stated
5. Sample size based on the assumptions as per above
	1. Reference the formula/sample size calculation approach, if standard binary, continuous or survival outcome formulae are not used. For a time-to-event outcome the number of events required should be stated.
	2. If any adjustments (e.g., allowance for loss to follow-up, multiple testing, etc.) that alter the required sample size are incorporated, they should also be specified, referenced, and justified along with the final sample size.
	3. For alternative designs, additional input should be stated and justified. For example, for a cluster RCT (or individually randomised RCTs with potential clustering) state the average cluster size and intra-cluster correlation coefficient(s). Variability in cluster size should be considered and, if necessary, the coefficient of variation should be incorporated into the sample size calculation. Justification for the values chosen should be given.
	4. Provide details of any assessment of the sensitivity of the sample size to the inputs used.

***Additional item for grant application and trial protocol***1. Underlying basis used for specifying the target difference (an *important* and/or *realistic* difference)
2. Explain the choice of target difference – specify and reference any formal method used or relevant previous research

***Additional item for trial results paper*** 1. Reference the trial protocol
 | ***Page and line numbers where item is reported*** |

|  |
| --- |
| **Box 3 Methods that can be used to inform the choice of the target difference*****Methods that inform what is an important difference****Anchor:* The outcome of interest can be ‘‘anchored’’ by using either a patient’s or health professional’s judgement to define what an important difference is. This may be achieved by comparing a patient’s health before and after treatment and then linking this change to participants who showed improvement/deterioration using a more familiar outcome (for which either patients or health professionals more readily agree on what amount of change constitutes an important difference).. Contrasts between patients (e.g., individuals with varying severity of a disease) can also be used to determine a meaningful difference.*Distribution:* Approaches that determine a value based upon distributional variation. A common approach is to use a value that is larger than the inherent imprecision in the measurement and therefore likely to represent a minimal level needed for a noticeable difference.*Health economic:* Approaches that use principles of economic evaluation. These compare cost with health outcomes, and define a threshold value for the cost of a unit of health effect that a decision-maker is willing to pay, to estimate the overall incremental net benefit of one treatment versus the comparator. A study can be powered to exclude a zero incremental net benefit at a desired statistical significance and power. A radically different approach is a (Bayesian) decision-theoretic value of information analysis which compares the added value with the added cost of the marginal observation, thus avoiding the need to specify a target difference. *Standardised effect size:* The magnitude of the effect on a standardised scale defines the value of the difference. For a continuous outcome, the standardised difference (most commonly expressed as Cohen’s d ‘‘effect size’’, the mean difference dividing by the standard deviation) can be used. Cohen’s cutoffs of 0.2, 0.5, and 0.8 for small, medium, and large effects, respectively, are often used. Thus a ‘‘medium’’ effect corresponds simply to a change in the outcome of 0.5 SDs. When measuring a binary or survival (time-to-event) outcome alternative metrics (e.g., an odds, risk, or hazard ratio) can be utilised in a similar manner, though no widely recognised cut-points exist. Cohen’s cut-points approximate odds ratios of 1.44, 2.48, and 4.27, respectively.32 Corresponding risk ratio values vary according to the control group event proportion.***Methods that inform what is a realistic difference****Pilot study:* A pilot (or preliminary) study may be carried out where there is little evidence, or even experience, to guide expectations and determine an appropriate target difference for the trial. In a similar manner, a Phase 2 study could be used to inform a Phase 3 study though this would need to take account of methodological differences (e.g. inclusion criteria and outcomes) that should be reflected in specification of the target difference.***Methods that inform what is an important and/or a realistic difference****Opinion-seeking:* The target difference can be based on opinions elicited from health professionals, patients, or others. Possible approaches include forming a panel of experts, surveying the membership of a professional or patient body, or interviewing individuals. This elicitation process can be explicitly framed within a trial context. *Review of evidence base:* The target difference can be derived from current evidence on the research question. Ideally, this would be from a systematic review or meta-analysis of RCTs. In the absence of randomised evidence, evidence from observational studies could be used in a similar manner.  |

**REFERENCES**

1. Altman D, Schulz K, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663 - 94.

2. Cook J, Hislop J, Adewuyi T, et al. Assessing methods to specify the targeted difference for a randomised controlled trial - DELTA (Difference ELicitation in TriAls) review. *Health technology assessment* 2014;18:28.

3. Hislop J, Adewuyi TE, Vale LD, et al. Methods for specifying the target difference in a randomised controlled trial: the Difference ELicitation in TriAls (DELTA) systematic review. *PLoS medicine* 2014;11(5)(e1001645) doi: 10.1371/journal.pmed.1001645

4. Cook JA, Hislop JM, Altman DG, et al. Use of methods for specifying the target difference in randomised controlled trial sample size calculations: Two surveys of trialists' practice. *Clinical trials* 2014;11(3):300-08. doi: 10.1177/1740774514521907

5. Cook JA, Julious SA, Sones W, et al. Choosing the target difference (“effect size”) for a randomised controlled trial - DELTA2 guidance 2018 [Available from: <https://drive.google.com/file/d/1QV_a7AKh9UYOaw6k0dkreHd8eXADkRqK/view> accessed 1/06/2018 2018.

6. Sones W, Julious SA, Rothwell JC, et al. Choosing the target difference (“effect size”) for a randomised controlled trial – the development of the DELTA2 guidance *Trials [submitted]* 2018

7. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *Bmj* 1999;319(7211):670-4. [published Online First: 1999/09/10]

8. Phillips A, Abellan-Andres J, Soren A, et al. Estimands: discussion points from the PSI estimands and sensitivity expert group. *Pharmaceutical statistics* 2017;16(1):6-11. doi: 10.1002/pst.1745 [published Online First: 2016/03/22]

9. Rosenkranz G. Estimands-new statistical principle or the emperor's new clothes? *Pharmaceutical statistics* 2017;16(1):4-5. doi: 10.1002/pst.1792 [published Online First: 2016/12/15]

10. Committee for Human Medicinal Products. ICH E9 (R1) addendum on estimands and Sensitivity Analysis in Clinical Trials to the guideline on statistical principles for clinical trials EMA/CHMP/ICH/436221/2017, 2017:1-23.

11. Akacha M, Bretz F, Ruberg S. Estimands in clinical trials - broadening the perspective. *Statistics in medicine* 2017;36(1):5-19. doi: 10.1002/sim.7033 [published Online First: 2016/07/21]

12. National Institute for Health Research. Involve 2017 [Available from: <http://www.invo.org.uk/> accessed 4/5/2017.

13. Chan KB, Man-Son-Hing M, Molnar FJ, et al. How well is the clinical importance of study results reported? An assessment of randomized controlled trials. *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne* 2001;165(9):1197-202. [published Online First: 2001/11/15]

14. Schulz KF, Grimes DA. Sample size calculations in randomised trials: mandatory and mystical. *Lancet* 2005;365(9467):1348-53. doi: 10.1016/s0140-6736(05)61034-3 [published Online First: 2005/04/13]

15. Senn S. Controversies concerning randomization and additivity in clinical trials. *Statistics in medicine* 2004;23(24):3729-53. doi: 10.1002/sim.2074

16. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. 1st ed. Chicester: John Wiley & Sons 2004.

17. Goodman SN, Berlin JA. The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Ann Intern Med* 1994;121(3):200-6. [published Online First: 1994/08/01]

18. Charles P, Giraudeau B, Dechartres A, et al. Reporting of sample size calculation in randomised controlled trials: review. *Bmj* 2009;338(b1732) doi: 10.1136/bmj.b1732

19. Bland JM. The tyranny of power: is there a better way to calculate sample size? *Bmj* 2009;339(b3985) doi: 10.1136/bmj.b3985 [published Online First: 2009/10/08]

20. Stallard N, Miller F, Day S, et al. Determination of the optimal sample size for a clinical trial accounting for the population size. *Biometrical journal Biometrische Zeitschrift* 2016;59(4):609-25. doi: 10.1002/bimj.201500228 [published Online First: 2016/05/18]

21. Pezeshk H. Bayesian techniques for sample size determination in clinical trials: a short review. *Statistical methods in medical research* 2003;12(6):489-504. [published Online First: 2003/12/05]

22. Claxton K. The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies. *Journal of health economics* 1999;18(3):341-64. [published Online First: 1999/10/28]

23. Julious S. Sample sizes for clinical trials: Chapman and Hall/CRC Press, Boca Raton, FL 2010.

24. Hellum C, Johnsen L, Storheim K, et al. Surgery with disc prosthesis versus rehabilitation in patients with low back pain and degenerative disc: two year follow-up of randomised study. *Bmj* 2011;342:d2786.

25. White PD, Goldsmith KA, Johnson AL, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. *Lancet* 2011;377(9768):823-36. doi: 10.1016/s0140-6736(11)60096-2 [published Online First: 2011/02/22]

26. Copay A, Subach B, Glassman S, et al. Understanding the minimum clinically important difference: a review of concepts and methods. *The spine journal : official journal of the North American Spine Society* 2007;7:541 - 46.

27. Wells G, Beaton D, Shea B, et al. Minimal clinically important differences: Review of methods. *J Rheumatol* 2001;28:406 - 12.

28. Beaton D, Boers M, Wells G. Many faces of the minimal clinically important difference (MICD): A literature review and directions for future research. *Curr Opin Rheumatol* 2002;14:109 - 14.

29. Fayers P, Cuschieri A, Fielding J, et al. Sample size calculation for clinical trials: the impact of clinician beliefs. *British journal of cancer* 2000;82:213 - 19.

30. Cook JA, Hislop J, Adewuyi TE, et al. Assessing methods to specify the target difference for a randomised controlled trial: DELTA (Difference ELicitation in TriAls) review. *Health technology assessment* 2014;18(28):v-vi, 1-175. doi: 10.3310/hta18280

31. Cook J, Hislop J, Altman D, et al. Specifying the target difference in the primary outcome for a randomised controlled trial: guidance for researchers. *Trials* 2015;16(12)

32. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in medicine* 2000;19(22):3127-31. [published Online First: 2000/12/13]