Practical help for specifying the target difference in sample size calculations for RCTs: the DELTA² five-stage study, including a workshop

Jonathan A Cook,[1]* Steven A Julious,[2] William Sones,[1] Lisa V Hampson,[3] Catherine Hewitt,[4] Jesse A Berlin,[5] Deborah Ashby,[6] Richard Emsley,[7] Dean A Fergusson,[8] Stephen J Walters,[2] Edward CF Wilson,[9,10] Graeme MacLennan,[11] Nigel Stallard,[12] Joanne C Rothwell,[2] Martin Bland,[13] Louise Brown,[14] Craig R Ramsay,[15] Andrew Cook,[16] David Armstrong,[17] Douglas Altman[1†] and Luke D Vale[18]

[1]Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK
[2]Medical Statistics Group, School of Health and Related Research, University of Sheffield, Sheffield, UK
[3]Statistical Methodology and Consulting, Novartis Pharma AG, Basel, Switzerland
[4]York Trials Unit, Department of Health Sciences, University of York, York, UK
[5]Johnson & Johnson, Titusville, NJ, USA
[6]Imperial Clinical Trials Unit, Imperial College London, London, UK
[7]Department of Biostatistics and Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[8]Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, ON, Canada
[9]Cambridge Centre for Health Services Research, Cambridge Clinical Trials Unit University of Cambridge, Cambridge, UK
[10]Health Economics Group, Norwich Medical School, University of East Anglia, Norwich, UK
[11]Centre for Healthcare Randomised Trials, University of Aberdeen, Aberdeen, UK
[12]Warwick Medical School, Statistics and Epidemiology, University of Warwick, Coventry, UK
[13]Department of Health Sciences, University of York, York, UK
[14]MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, UK
[15]Health Services Research Unit, University of Aberdeen, Aberdeen, UK
[16]Wessex Institute, University of Southampton, Southampton, UK

[17]School of Population Health and Environmental Sciences, King's College London, London, UK
[18]Health Economics Group, Institute of Health & Society, Newcastle University, Newcastle upon Tyne, UK

*Corresponding author jonathan.cook@ndorms.ox.ac.uk
†In memoriam

Declared competing interests of authors: Lisa V Hampson is an employee of Novartis Pharma AG (Basel, Switzerland) and reports grants from the Medical Research Council (MRC). Catherine Hewitt is a member of the National Institute for Health Research (NIHR) Health Technology Assessment (HTA) Commissioning Board since 2015. Jesse A Berlin is an employee of Johnson & Johnson (New Brunswick, NJ, USA) and holds shares in this company. Richard Emsley is a member of the NIHR HTA Clinical Trials Board since 2018. Deborah Ashby is a member of the HTA Commissioning Board, HTA Funding Boards Policy Group, HTA Mental Psychological and Occupational Health Methods Group, HTA Prioritisation Group and the HTA Remit and Competitiveness Group from January 2016 to December 2018. Stephen J Walters declares his department has contracts and/or research grants with the Department of Health and Social Care, NIHR, MRC and the National Institute for Health and Care Excellence. He also declares book royalties from John Wiley & Sons, Inc. (Hoboken, NJ, USA), as well as a grant from the MRC and personal fees for external examining. Louise Brown is a member of the NIHR Efficacy and Mechanism Evaluation Board since 2014. Craig R Ramsay is a member of the NIHR HTA General Board since 2017. Andrew Cook is a member of the NIHR HTA Interventional Procedures Methods Group, HTA Intellectual Property Panel, HTA Prioritisation Group, Public Health Research (PHR) Research Funding Board, Public Health Research Prioritisation Group and the PHR Programme Advisory Board.

Published October 2019
DOI: 10.3310/hta23600

Scientific summary

The DELTA^2 five-stage study, including a workshop

Health Technology Assessment 2019; Vol. 23: No. 60
DOI: 10.3310/hta23600

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

This report summarises the Difference ELicitation in TriAls[2] (DELTA[2]) advice and recommendations for researchers and funder representatives on specifying the target difference and undertaking a sample size calculation for a randomised controlled trial. Details of the work carried out to inform the development of the document are also provided in the report. A summary of the key topics and recommendations for practice and reporting are provided below.

Specifying the target difference for a randomised controlled trial

The randomised controlled trial is widely considered the gold standard study for comparing the effectiveness of health interventions. Central to its design is a calculation of the number of participants needed – the sample size. This provides reassurance that the study will be able to achieve its primary aim. It is typically done by specifying the magnitude of the difference between the intervention effects in the key (primary) outcome for the population of interest that can reliably be detected for a given sample size. This difference is called the study's 'target difference' and should be appropriate for the primary estimand of interest (i.e. the combination of population, outcome and intervention effects), as determined by the primary aim of the study.

There are two main bases for specifying a target difference: (1) a difference that is considered to be important to one or more stakeholder groups (e.g. patients); and/or (2) a difference that is realistic (plausible), based on existing evidence and/or expert opinion. Seven broad types of methods can be used to justify the choice of a particular value as the target difference: (1) anchor, (2) distribution, (3) health economic, (4) opinion-seeking, (5) pilot study, (6) review of the evidence base and (7) standardised effect size.

Different statistical and health economic approaches can be taken to justify the sample size, but the general principles are mostly the same. An exception is the relatively new technique of value of information analysis, which seeks to explicitly incorporate the opportunity cost of conducting research. In this case, the appropriate sample size is one that maximises the return on investment in the trial, dispensing with the need to define a target difference. The use of alternative approaches is currently limited, with the conventional (Neyman–Pearson) approach the most commonly used.

To aid those new to the topic and to encourage better practice regarding the specification of the target difference for a randomised controlled trial, the following recommendations are made when the conventional approach to the sample size calculation is used.

- Begin by searching for relevant literature to inform the specification of the target difference. Relevant literature can:
  - relate to a candidate primary outcome and/or the comparison of interest
  - inform what is an important and/or realistic difference for that outcome, comparison and population (estimand of interest).
- Candidate primary outcomes should be considered in turn and the corresponding sample size explored. When multiple candidate outcomes are considered, the choice of primary outcome and target difference should be based on consideration of the views of relevant stakeholder groups (e.g. patients), as well as the practicality of undertaking such a study and 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, although this is not always practical.

- 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.
- 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.
- 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.
- When additional research is needed to inform what would be an important difference to one or more stakeholder groups (e.g. patients), the anchor and opinion-seeking methods are to be favoured. The distribution method should not be used. Specifying the target difference based solely on a standardised effect size approach should be considered a last resort, although it may be helpful as a secondary approach.
- When additional research is needed to inform what would be a realistic difference, the opinion-seeking and 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.
- Use existing studies to inform the value of key nuisance parameters that are part of the sample size calculation. For example, a pilot trial can be used to inform the choice of standard deviation value for a continuous outcome or the control group proportion for a binary outcome, along with other relevant inputs, such as the number of missing outcome data.
- Sensitivity analyses that 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.
- Specification of the sample size calculation, including the target difference, should be reported in accordance with the recommendations for reporting items (see *Recommended core reporting items*) when preparing key trial documents (grant applications, protocols and result manuscripts).

Recommended core reporting items

A set of core items should be reported in all key trial documents (protocols, grant applications and main results papers) to ensure reproducibility of the sample size calculation. Recommended core reporting items when the conventional sample size approach has been used are as follows:

- Primary outcome (and any other outcome on which the calculation is based) –
    - If a primary outcome is not used as the basis for the sample size calculation, state why.
- Statistical significance level and power.
- Express the target difference according to outcome type –
    - 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.
    - Continuous: state the target mean difference on the natural scale, the common standard deviation and the standardised effect size (mean difference divided by the standard deviation).
    - Time to event: state the target difference as an absolute and/or relative difference and provide the control group event proportion; the planned length of follow-up; 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.
- Allocation ratio –
    - If an unequal ratio is used, the reason for this should be stated.

- Sample size based on the assumptions as per above.
    - 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.
    - If any adjustments (e.g. allowance for loss to follow-up, multiple testing) that alter the required sample size are incorporated, they should also be specified, referenced and justified, along with the final sample size.
    - For alternative designs, any additional inputs should be stated and justified. For example, for a cluster randomised controlled trial (or individually randomised trials with potential clustering), state the average cluster size and intracluster 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.
    - Provide details of any assessment of the sensitivity of the sample size to the inputs used.

Trial results papers should always reference the trial protocol. Additional items to give further explanation of the rationale should be provided when space allows (e.g. grant applications and trial protocols). When the calculation deviates from the conventional approach, whether by research question or statistical framework, this should be clearly specified. The reporting items would correspondingly need appropriate modification.

Funding

Funded by the Medical Research Council (MRC) UK and National Institute for Health Research as part of the MRC–National Institute for Health Research Methodology Research programme.

Health Technology Assessment

HTA MRC MSAW

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

Impact factor: 3.819

Health Technology Assessment is indexed in MEDLINE, CINAHL, EMBASE, The Cochrane Library and the Clarivate Analytics Science Citation Index.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta. Print-on-demand copies can be purchased from the report pages of the NIHR Journals Library website: www.journalslibrary.nihr.ac.uk

Criteria for inclusion in the *Health Technology Assessment* journal

Reports are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme or, commissioned/managed through the MRC–NIHR Methodology Research Programme (MRP), and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This report

This issue of the Health Technology Assessment journal series contains a project commissioned by the MRC–NIHR Methodology Research Programme (MRP). MRP aims to improve efficiency, quality and impact across the entire spectrum of biomedical and health-related research. In addition to the MRC and NIHR funding partners, MRP takes into account the needs of other stakeholders including the devolved administrations, industry R&D, and regulatory/advisory agencies and other public bodies. MRP supports investigator-led methodology research from across the UK that maximises benefits for researchers, patients and the general population – improving the methods available to ensure health research, decisions and policy are built on the best possible evidence.

To improve availability and uptake of methodological innovation, MRC and NIHR jointly supported a series of workshops to develop guidance in specified areas of methodological controversy or uncertainty (Methodology State-of-the-Art Workshop Programme). Workshops were commissioned by open calls for applications led by UK-based researchers. Workshop outputs are incorporated into this report, and MRC and NIHR endorse the methodological recommendations as state-of-the-art guidance at time of publication.

The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

This report presents independent research funded under a MRC–NIHR partnership. The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the MRC, NETSCC, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the MRC, NETSCC, the HTA programme or the Department of Health and Social Care.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk).