Introduction: We are seeking to determine: a) optimal approaches to address statistical multiplicity in pragmatic RCTs, and b) whether approaches should vary according to the taxonomy of the situation and context of research questions. The first step is to collate information from three sources:
a)Systematic literature review around opinions, guidelines and methods used
b)Survey of practice within UK based Clinical Trial Units (CTUs)
c)Survey of the methods used in pragmatic RCTs recently published in high impact medical journals
Methods: A key objective is the creation of a conceptual framework to synthesise the information obtained from the three sources. We are extracting data from relevant papers into nVivo software in order to generate themes and visualise findings graphically, e.g. comparison diagrams and concept maps. These findings will also be used to inform the design of both surveys, which will include factual questions about specific approaches and open-ended questions to capture alternative methods.
Provisional results: Provisional systematic review findings suggest varying opinions/guidelines on the approach to multiplicity, and that methods used range from simple Bonferroni adjustment through to more complicated procedures, e.g. hierarchical/gatekeeping procedures. Once results from all three sources are available the synthesised findings will be triangulated to inform the rest of the research; the first step of which is planned to be empirical studies comparing probabilities of Type I and II errors when applying different methods to different multiplicity situations.
Potential relevance and impact: This research will provide guidance on recommended approaches to address multiplicity for methodologists designing pragmatic RCTs and health professionals interpreting RCT findings. The overall results have potential to be applied to most RCTs commissioned by the NIHR, hence influencing evidence-based changes in practice in the UK NHS across many disease areas.
This research is funded by the NIHR and the NIHR disclaimer applies.

PS3D

- O2 Assessing the impact of early stopping on systematic reviews: Recommendations for interpreting guidelines

Ian Marschner[1,2], Lisa Askie[1,2], Manjula Schou[3]
[1]NHMRC Clinical Trials Centre, Sydney, Australia; [2]University of Sydney, Sydney, Australia; [3]Macquarie University, Sydney, Australia
Trials 2019, **20(Suppl 1):**PS3D

Introduction: The CONSORT Statement says that early stopping of a clinical trial weakens the conclusions from the trial. The GRADE guidelines go further, saying that early stopping is a study limitation carrying the risk of bias, and recommending systematic reviewers conduct sensitivity analyses omitting such studies. Our aim is to assess methodologies for conducting these sensitivity analyses and to make recommendations about interpretation of the guidelines.
Methods: We began by reviewing and summarising the range of possible impacts of early stopping on treatment effect estimation in single studies and meta-analyses. We then used simulation studies to evaluate the performance of various approaches to meta-analysis when early stopping is present. Our primary focus was early stopping for benefit and we investigated the performance of meta-analyses where treatment effect estimates were adjusted for the interim monitoring using the statistical method of conditional maximum likelihood estimation.
Results: Early stopping at the first interim analysis leads to overestimation of treatment effects but at subsequent interim analyses the effect is unpredictable, and may lead to overestimation, underestimation or even unbiased estimation. Interim monitoring that does not stop the study early leads to underestimation. Importantly, these effects do not translate into biased meta-analyses, because the underestimation and overestimation balance each other in multiple studies. In contrast, exclusion of studies stopped early always leads to underestimation in meta-analyses. However, if treatment effect estimates are adjusted for the interim monitoring, prior to conducting the meta-analysis, the underestimation is rectified.
Discussion: We recommend against sensitivity analyses that simply exclude studies that stopped early from meta-analyses. If a sensitivity analysis is conducted in accordance with GRADE, we recommend that treatment effect estimates are first adjusted for the interim monitoring. To facilitate this, we recommend specific information that should be reported in adhering to the CONSORT reporting standards on early stopping.

PS3D

- O3 Reporting of methodological aspects of randomised trials: 1996-2016; has it changed over time?

Shona Fielding, Ruiyi Long, Neil Scott
University Of Aberdeen, Aberdeen, United Kingdom
Trials 2019, **20(Suppl 1):**PS3D

Introduction: The Consolidated Standards of Reporting Trials (CONSORT) statement was introduced in 1996 and revised in 2010. It provides a checklist for reporting details of the study design/results of randomised controlled trials (RCTs). We aimed to survey RCTs published in 1996, 2006 and 2016 to determine whether reporting standards of key methodological details had changed over time.
Method: We used MEDLINE to obtain a random sample of published RCT articles from each year. There was no restriction on journal. Titles and abstracts were screened, and full text of potentially eligible studies assessed for inclusion. We collected information on the journal, number of treatment arms, whether a CONSORT flow diagram and power calculation were included and how randomisation methods and baseline data were reported.
Result: 252/603 studies were considered eligible for inclusion. Just over a third provided a CONSORT flow diagram. A table of baseline characteristics was provided in 83% of studies, but over half of these also provided p-values for baseline data. Details of a power calculation were included for 47% of studies. The method used to generate the random sequence and method of allocation concealment was unclear/not reported in 70% and 65% of studies respectively. Over time, the proportions including CONSORT diagrams, power calculations and baseline tables increased. Reporting of randomisation methods also improved; permuted blocks and sealed envelopes remained the most common methods for sequence generation and concealment.
Discussion: Our survey was not restricted to high impact journals and showed that reporting standards have improved between 1996 and 2016, but despite the existence of CONSORT, there is still poor reporting of key methodological aspects of RCTs. Notably, over half of articles reported significance tests for baseline characteristics, contrary to current recommendations, and the method of randomisation was either unclear or not reported in over half of the studies.

PS3D

- O4 How well are binary outcomes analysed and the findings reported? – A systematic review of randomised trials

Ines Rombach[1,2,3], Nicholas Peckham[1,2,3], Ruth Knight[1,2,3], Jamie Stokes[1,2,3], Jonathan A Cook[1,2,3]
[1]Oxford Clinical Trials Unit, Oxford, UK; [2]Centre for Statistics in Medicine, Oxford, UK; [3]Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, Oxford, UK
Trials 2019, **20(Suppl 1):**PS3D

Introduction: Randomised clinical trials (RCTs) need to be reported in a way that enables robust and reliable interpretation of their results. Here we review how well current publications of RCTs report binary primary outcomes and appropriate sensitivity analysis for missing data.
Numerous different statistical approaches exist for the analysis of binary data, some of which have been shown to be suboptimal. Guidelines, including the Consolidated Standards of Reporting Trials (CONSORT), stipulate that analyses of binary outcomes should present estimates for both absolute and relative effects, together with confidence intervals. Where some primary outcome data are missing, sensitivity analysis should investigate the effect of changes

to assumed missing data mechanism on the trial conclusions. It is uncertain what is typically done.
Methods: We searched MEDLINE for primary publications of RCTs published in January 2019, and identified those that reported a binary primary endpoint. Data to be extracted include the study size, loss to follow-up by trial arm, statistical analysis methods (principal and secondary) for the primary endpoint, together with whether or not absolute as well as relative effects and their confidence intervals were reported. Details of sensitivity analyses performed will be extracted, and reporting of the results will be considered
Timing of Potential Results: The search and study assessment process is in progress and data extraction and synthesis will take place over the summer. Around 200 trials are anticipated to be included.
Potential Relevance & Impact: Current practice on how RCTs with binary primary endpoints are analysed, and how well the results are presented will be summarised. Presentation of both relative and absolute effects, as well as the use of appropriate sensitivity analysis are assessed. This will be useful to both authors and reviewers of RCT reports and contribute to improving reporting standards, as well as identifying potential areas for improvement.

PS3D

- O5 Overestimation of Event Rate and Target Difference among Randomized Clinical in sample size calculations Trials: a cross-sectional survey review

Tao Chen[1], Chao Li[2], Yang Wang[3], James Dodd[1], Victoria Cornelius[4], Duolao Wang[1]
[1]Liverpool School Of Tropical Medicine, Liverpool, United Kingdom; [2]Xi'an Jiaotong University Health Science Centre, Xi'an, China; [3]Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing , China; [4]Imperial College London, LONDON, United Kingdom
Trials 2019, **20(Suppl 1):**PS3D

Background: When designing a randomized clinical trial (RCT), unlike power and alpha which can be set at conventionally accepted values (usually 90% or 5%, respectively), assumptions about the event rate of the primary endpoint and the targeted difference (or minimal clinically important difference) for the sample size calculation are often based on prior knowledge from empirical studies or expert opinion. Overestimation of event rates and target difference can have a detrimental impact on the power of an RCT. In this study, we aim to systematically investigate the prevalence and extent of overestimation of event rates and targeted difference. We will then examine their impact on trial conclusions.
Methods: We are planning to perform a cross-sectional survey review of phase II-III RCTs published in seven high impact medical journals from January 2015 to January 2019.
Timing of Potential Results: This study is part of an ongoing clinical trials survey. It will include approximately 200 two-arm, parallel and superiority trials with a single binary outcome. In a pilot review of the first 68 RCTs, we have found that 34 (50.0%) had an overestimated primary event rate (i.e. smaller than the lower limit of the 95% CI of observed event rate) and 31 (45.6%) had a larger than targeted difference (i.e. outside the 95% CI of the actual treatment difference). However, among those 41 studies either having overestimation of event rate or target difference, 10 of them drew a positive conclusion.
Potential Relevance & Impact: We think that many trials may be underpowered due to higher event rates or aiming to detect unrealistic treatment effects. If this is the case the results from these studies should be interpreted with caution and not in isolation. This is particularly important for trials with positive conclusion.

PS4A

- O1 Internal pilots in clinical trials: Current practice in design and assessment

Anna Rosala-Hallas[1], Carrol Gamble[1], Jane Blazeby[2], Paula R Williamson[3]
[1]Clinical Trials Research Centre, University of Liverpool, a member of the Liverpool Health Partners, Liverpool, United Kingdom; [2]Medical Research Council ConDuCT II Hub for Trials Methodology Research, NIHR Bristol Biomedical Research Centres, Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom; [3]North West Hub for Trials Methodology Research/Clinical Trials Research Centre, University of Liverpool, a member of the Liverpool Health Partners, Liverpool, United Kingdom
Trials 2019, **20(Suppl 1):**PS4A

Introduction: Running an internal pilot can help to optimise methods for a trial. The literature provides recommendations for the design of internal pilots, however, information is lacking regarding the designs used and the process of review against progression criteria. The aim of this research is to provide an overview of current practice.
Methods: A cohort of clinical trials, extracted in 2017, comprised those with an internal pilot having undergone the progression review stage, funded by the National Institute for Health Research Health Technology Assessment programme. Data about the design and assessment of the internal pilots were abstracted from: project descriptions; summary of changes from the first stage application; funder/reviewer feedback about the full application; funder monitoring notes; progress report history and trial protocols.
Results: Fifty-seven studies were reviewed. The majority of internal pilots were first proposed in the initial stages. The Trial Steering Committee was mostly commonly involved in the review process, alongside the funder. Progression criteria included: target number for recruitment, rate of randomisation, retention/primary outcome ascertainment rate, rate of treatment adherence and consent rate. All studies but one continued to the main trial, however a quarter did not strictly meet their progression criteria. Actions taken for studies which did not meet their progression criteria included a second review, recovery plan and close monitoring. Changes were made to the design of the main trial for 25% of the studies; these were primarily to do with conduct of recruitment.
Discussion: We provide insight into the process of designing and assessing internal pilot trials. Progression criteria are sometimes not met; however, committees involved in the reviewing process will generally support continuation to the main trial, usually accompanied by a second review or close monitoring. We make recommendations to optimise the decision-making process.

PS4A

- O2 External Pilot and Feasibility Studies: Past, Present and Future Challenges

Lehana THABANE[1], Sandra Eldridge[2], Christine Bond[3], Claire Chan[2], Gillian Lancaster[4], Michael Campbell[5], Sally Hopewell[6]
[1]McMaster University, Hamilton, Canada; [2]Queen Mary University of London, London, UK; [3]University of Aberdeen, Aberdeen, UK; [4]Keele University, Keele, UK; [5]University of Sheffield, Sheffield, UK; [6]University of Oxford, Oxford, UK
Trials 2019, **20(Suppl 1):**PS4A

External pilot and feasibility (PAF) trials are an essential part of trial preparation, particularly for the planning of complex interventions. However, they rarely published, and the ones that are