Reporting of multi-arm parallel group randomized trials: extension of the CONSORT 2010 statement

Edmund Juszczak[1] MSc, Douglas G Altman[2] DSc (†), Sally Hopewell[2] DPhil, Kenneth Schulz[3] PhD

1. NPEU Clinical Trials Unit, National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK

2. Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK (†Deceased)

3. FHI 360, 359 Blackwell Street, Suite 200, Durham, North Carolina, USA and The University of North Carolina at Chapel Hill, School of Medicine, Chapel Hill, North Carolina, USA

Corresponding author details: Associate Professor Edmund Juszczak, NPEU Clinical Trials Unit, National Perinatal Epidemiology Unit, Nuffield Department of Population Health, **University of Oxford**, Old Road Campus, Headington, Oxford, OX3 7LF, UK. Direct telephone +44 - (0)1865 289743; Jenny Shilton Osborne (PA to Edmund Juszczak) +44 - (0)1865 289748; Email: ed.juszczak@npeu.ox.ac.uk

Word count: 5,597 (not including title, key points, abstract, acknowledgment, references, tables, and figure legends)

Key Points

Question

What additional information should be provided when reporting a multi-arm randomized trial?

Findings

This reporting guideline is an extension of the 2010 version of the Consolidated Standards of Reporting Trials (CONSORT) Statement. Ten CONSORT items have been modified, and examples of good reporting and an accompanying explanation for each extension item is provided.

Meaning

The guideline checklist can facilitate transparent reporting of multi-arm randomized trials, and may help assist evaluations of rigor and reproducibility, enhance understanding of the methodology, and make the results more useful for clinicians, journal editors, reviewers, guideline authors, and funders.

Abstract

IMPORTANCE: The quality of reporting of randomized clinical trials (RCTs) is sub-optimal. In an era in which the need for greater research transparency is paramount, poor reporting hinders assessment of the reliability and validity of trial findings. The Consolidated Standards of Reporting Trials (CONSORT) Statement was developed to improve the reporting of RCTs, but the primary focus was on parallel group trials with two treatment groups.

Multi-arm trials that use a parallel group design but have three or more groups are relatively common. The quality of reporting of multi-arm trials varies substantially, making judgements and interpretation difficult. While the majority of the elements of the CONSORT Statement apply equally to multi-arm trials, some elements need adaptation, and in some cases, additional issues need to be clarified.

OBJECTIVE: To present an extension to the 2010 version of the CONSORT Statement for reporting multi-arm trials to facilitate the reporting of such trials.

DESIGN: A guideline writing group, which included all four authors, formed following the CONSORT group meeting in 2014. The authors met face-to-face and by teleconference regularly on multiple occasions between 2014 and 2018 to develop and revise the checklist and the accompanying text, with additional discussions by email.

No Delphi process was conducted. A draft manuscript was circulated for review to the wider CONSORT group (n=36) plus five other selected individuals known for their specialist knowledge in clinical trials. Extensive feedback was received from 14 individuals and after detailed consideration of their comments, a final revised version of the extension was prepared.

FINDINGS: This CONSORT extension for multi-arm trials extends ten items on the CONSORT 2010 checklist and provides examples of good reporting and a rationale for the importance of each extension item. Key recommendations are that multi-arm trials should be identified as such and require clear objectives and hypotheses referring to all of the treatment groups. Primary treatment comparisons should be identified in advance and authors should report the comparisons resulting

from multiple groups completely and transparently, planned and unplanned. If statistical adjustments for multiplicity are applied, the rationale and method used should be described. CONCLUSIONS AND RELEVANCE: This extension of the CONSORT 2010 Statement provides specific guidance for the reporting of multi-arm parallel group RCTs.

Introduction

Multi-arm randomized controlled trials can take a number of different forms but are typically a combination of elements including multiple active interventions, combination(s) of active interventions, different doses (or regimens) of an intervention, placebo, and no active intervention, or treatment as usual. These elements can be combined in various ways resulting in numerous possible trial structures. For example, in a three-arm trial of A1 versus A2 versus A3, this could represent an evaluation of different doses of the same active intervention. Alternatively, A1 versus B1 versus C1 could represent an evaluation of two different active interventions versus a control. A1 versus A2 versus B1 could represent an evaluation of two different doses of an active intervention versus another active intervention.

Evaluating more than one new intervention concurrently increases the chances of finding an effective intervention (1). The corresponding increase in efficiency using a multi-arm design, compared to performing sequential two-arm trials, should result in lower cost due to better use of resources. In most cases, sharing a control arm reduces the sample size relative to performing separate two-arm trials (2). By typically offering participants a higher probability of being allocated to a new intervention, this may result in a greater proportion of eligible people enrolling. Some multi-arm trials in oncology have recruited more quickly than comparable two-arm trials (1). The argument against multi-arm trials primarily centres on statistical power since published trials are often compromised on this (3). Given a finite number of potential participants, adding additional treatment groups can further dilute already insufficient power.

Multi-arm trials are relatively common, a detailed review of all randomized trials indexed in PubMed published in one month in 2012 showed that 79% (1,062/1,351) were parallel group trials (4); of these, 14% (149/1,062) had 3 arms and 7% (76/1,062) had 4 or more arms.

In this article, we present guidance and an extension of the CONSORT checklist for the reporting of multi-arm trials, based on the 2010 version of the CONSORT Statement (5, 6). We provide illustrative examples and explanations for those items that differ from the main CONSORT checklist. We define multi-arm to mean randomized controlled trials that use a parallel group design but have three or more groups. Other multi-arm designs such as factorial, multi-arm multi-stage and adaptive trial designs raise rather different issues and will not be considered here.

Guidance development methods

Writing Group

The guideline writing group, which included all four authors, formed following a meeting of the CONSORT Group in 2014. The authors met face-to-face and by teleconference regularly on multiple occasions between 2014 and 2018 to develop and revise the checklist and accompanying examples and the accompanying text, with additional discussions by email.

Search strategy

To identify articles relevant to the methodology for multi-arm randomized trials, we undertook a scoping search of PubMed using the free text terms “multiarm”, “multi-arm” “multiple arm”, “multiple treatment” and “multiplicity” combined with the Publication Type term “Randomized Controlled Trial” as a topic, which identified 247 potential articles. One author (SH) assessed the titles and abstracts for relevance and as potentially relevant to this CONSORT extension. The search was supplemented with relevant articles from the personal collections of the authors and by searching the table of contents of books relevant to the methodology of clinical trials for information specific to the conduct and reporting of multi-arm trials.

Review and Refinement

No formal Delphi process was used in developing this CONSORT extension checklist; however, the draft manuscript was circulated in April 2017 for review to the wider CONSORT group, which included 36 individuals, plus five selected individuals known for their specialist knowledge in this area. Feedback was received from 14 individuals and after detailed consideration of their comments, a final revised version of the extension checklist and accompanying explanation was prepared.

Results

Checklist Items and Explanation

Table 1 shows the modified checklist; some items are extended to cover the reporting requirements relating to the multi-arm design, acknowledging the added complexity imposed by this design. Items requiring an extension from the CONSORT 2010 Statement are explained, with illustrative examples of good reporting. For items not mentioned below the advice is as for two-group, parallel randomized trials.

As all examples have been taken from published articles, it is inevitable that several do not display all the desirable elements of good reporting. Where this is the case, or where there might be ambiguity, the specific aspects of reporting which are addressed are identified. In some examples, text has been added in square brackets to explain the context. The CONSORT 2010 checklist for reporting the abstract of a randomized trial was reviewed. No separate checklist for abstracts is proposed, with the one proviso that authors report all of the objectives clearly and specify the number of treatment groups.

CONSORT checklist extension for multi-arm trials

Title and abstract

Item 1a

Standard CONSORT item: identification as a randomised trial in the title

Extension for multi-arm trials: identification as a multi-arm randomised trial in the title or an indication of the number of treatment groups that the participants were randomly assigned to

Examples

"HARMONY 3: 104-week randomized, double-blind, placebo- and active-controlled trial assessing the efficacy and safety of albiglutide compared with placebo, sitagliptin, and glimepiride in patients with type 2 diabetes taking metformin." (7)

"Efficacy of Oral Risperidone, Haloperidol, or Placebo for Symptoms of Delirium Among Patients in Palliative Care: A Randomized Clinical Trial." (8)

"Multiple Sclerosis-Secondary Progressive Multi-Arm Randomisation Trial (MS-SMART): a multiarm phase IIb randomised, double-blind, placebo-controlled clinical trial comparing the efficacy of three neuroprotective drugs in secondary progressive multiple sclerosis." (9)

Explanation

The ability to identify a report of a randomized trial in an electronic database depends largely on how it was indexed. Indexers may not classify a report as a randomized trial if the authors do not explicitly report this information (10). To help ensure that a study is appropriately indexed and easily identified, authors should use the word "randomized" in the title and indicate the number of arms (treatment groups) that the participants were randomly assigned to. This difficulty applies to multi-arm trials also. Article titles normally have a restricted word count, and so listing some or all of the

interventions is cumbersome and so adding the word multi-arm instead would be both economical and informative.

Introduction

Background and objectives

Item 2a

Standard CONSORT item: scientific background and explanation of rationale

Extension for multi-arm trials: rationale for using a multi-arm design

Example

"Many patients do not respond to monotherapy, and combinations of drugs are often recommended despite little evidence. Lithium plus valproate is often recommended after failure of first-line monotherapy. Should this combination have additive pharmacological effects and prove better than monotherapy, it could be an appropriate first-line therapy. We report here on BALANCE (Bipolar Affective disorder: Lithium/ANti-Convulsant Evaluation), a randomized trial that was designed to establish whether lithium plus valproate semisodium is better than monotherapy with either drug alone for prevention of relapse in bipolar I disorder." (11)

Explanation

When a trial compares two parallel groups it is evident that the aim is a comparison of those groups. With three or more 'intervention' arms, however, the intended main comparison or comparisons may not be clear. Since each intervention arm should be included only if it contributes to a specific research question, it follows that each arm should contribute to at least one pre-planned comparison. Authors should give a robust justification for using a multi-arm design and, in the

introduction to their article, say why they chose to investigate the interventions they studied and which specific comparisons were planned. In the situation, for example, where one of the planned interventions is a combination of two active interventions, authors might comment why they did not do a factorial trial. Typically, this 'incomplete' factorial design is used when it would be unethical to withhold active treatment from one group of patients.

Item 2b

Standard CONSORT item: specific objectives or hypotheses

Extension for multi-arm trials: specification of the research question referring to all of the treatment groups. Clear statement of all hypotheses to be tested and primary comparisons involved.

Example

Abstract (Objective): "To determine efficacy of risperidone or haloperidol relative to placebo in relieving target symptoms of delirium associated with distress among patients receiving palliative care."

Introduction: "The aim of this study was to determine if risperidone or haloperidol, given in addition to managing precipitants of delirium and providing individualized supportive nursing care, provides additional benefits in reducing target symptoms of delirium associated with distress when compared with placebo. The primary null hypothesis was that there was no difference between risperidone and placebo, and secondarily, no difference between haloperidol and placebo."(8)

Explanation

Eight possible analyses emanate from one three-arm trial (A, B, and C) of which most trials will include two or three (Box 1). The number of potential comparisons proliferates rapidly as the number of intervention groups increases – each group should appear in at least one comparison.

Thus, unless the intention is only to compare all groups at once (which is not a particularly sensible approach except perhaps for a dose-response study) there will be at least k–1 comparisons made in the analysis of a trial with k treatment arms. The maximum number of two group/paired comparisons is $k(k\text{-}1)/2$, e.g. for a four-arm trials there are six possible two-group comparisons.

Pre-specification of analyses is thus particularly important, and authors should report all the planned primary, secondary and exploratory comparisons. Otherwise, there is a major risk of highlighting and being misled by an observed difference without considering the large number of possible analyses. In all cases, and especially when many comparisons are planned, it is helpful to indicate the primary comparison(s). These comparisons should also feature in the explanation of the planned sample size (Item 7a). The planned comparisons may not be considered equally important. For example, one two-group comparison may be the primary focus of the trial. This is relevant when considering whether to make an adjustment for multiple comparisons. Alternatively, a hierarchical approach to hypothesis testing could prevent any issues with multiple comparisons (Item 12a). Some multi-arm trials combine a test of superiority with a test of non-inferiority. For example, Foa et al examined, among active military with posttraumatic stress disorder (PTSD), whether 10 sessions of prolonged exposure therapy (a trauma-focused cognitive behavioural therapy) delivered over 2 weeks (massed therapy) was more effective than minimal contact [control] and non-inferior to 10 sessions delivered over 8 weeks (spaced therapy) for reducing PTSD symptom severity (12).

Methods

Trial design

Item 3a

Standard CONSORT item: description of trial design (such as parallel, factorial) including allocation ratio

Extension for multi-arm trials: specification of the number of treatment groups

Examples

"In this pragmatic, open-label randomised trial, patients newly diagnosed with Parkinson's disease were randomly assigned (by telephone call to a central office; 1:1:1) between levodopa-sparing therapy (dopamine agonists or MAOBI [monoamine oxidase type B inhibitors]) and levodopa alone." (13)

"This was a phase 3, randomized, double-blind, placebo- and active-controlled parallel-group study that occurred between 17 February 2009 and 21 March 2013. Eligible patients were stratified by HbA1c level (<8.0% [<63.9 mmol/mol] vs. ≥8.0% [≥63.9 mmol/mol]), history of myocardial infarction (MI), and age (<65 vs. ≥65 years) and were randomly assigned (3:3:3:1) to receive, in addition to their background metformin, 1 of 4 treatments at baseline: albiglutide 30 mg, sitagliptin 100 mg, glimepiride 2 mg, or placebo. Matching placebos for albiglutide, sitagliptin, and glimepiride were used to maintain blinding to treatment."(7) An improvement on the reporting would be to explain why a 3:3:3:1 allocation was used.

Explanation

In terms of readability and understanding the design and rationale of a multi-arm trial, specification of the number of treatment groups is essential. Describing the allocation ratio offers insight and clarity, especially if an unequal allocation ratio is chosen, in which case an explanation is necessary.

Illustrating the structure and participant flow in a multi-arm trial will almost always reward the reader with valuable insight (an example demonstrating a trial structure and participant flow is shown in the eFigure in the Supplement) (7). Nevertheless, the presentation of the trial structure and participant flow in this example could still be improved in terms of the labelling (e.g. position of 'Follow-up' in diagram A), the absence of two arrows leading from the randomization box and the description of the information provided (e.g. what is meant by 'Terminated by sponsor' in diagram B?).

Item 3b

Standard CONSORT item: important changes to methods after trial commencement (such as eligibility criteria), with reasons

Extension for multi-arm trials: details of any treatment groups added or dropped (if relevant) with reasons and/or changes to allocation ratio

Example (where an arm was dropped)

"The original study was a multicentre, blinded, randomized, parallel-group trial in which patients were assigned to receive risperidone (Risperdal, Eisai), donepezil, or placebo for 12 weeks, after 4 weeks of psychosocial treatment. The target sample size was 285 people with Alzheimer's disease. Recruitment started in November 2003 but was suspended in March 2004, following the recommendation by the United Kingdom Committee for Safety of Medicines that risperidone and olanzapine not be used for the treatment of behavioral symptoms in dementia. The trial was restarted in July 2004 with a two-group design (donepezil and placebo), and recruitment ended in September 2005." (14)

Example (where an arm was added)

"A total of 1493 patients with schizophrenia were recruited at 57 U.S. sites and randomly assigned to receive olanzapine (7.5 to 30 mg per day), perphenazine (8 to 32 mg per day), quetiapine (200 to 800

mg per day), or risperidone (1.5 to 6.0 mg per day) for up to 18 months. Ziprasidone (40 to 160 mg per day) was included after its approval by the Food and Drug Administration. The primary aim was to delineate differences in the overall effectiveness of these five treatments."(15)

Explanation

If treatment arms are added or dropped, this affects the number of participants available for an unbiased and valid comparison i.e. only those participants randomized concurrently should be compared. In the example above where a treatment arm was dropped, the allocation ratio went from 1:1:1 to 1:1 (evident from the participant flow chart and results tables), so the probability of receiving one of the interventions changed from 0.33 to 0.5, but nevertheless randomization continued with roughly equal probability of receiving either intervention. In the example of adding an arm, the allocation ratio was not explicitly ever mentioned.

This item relates to a conventional multi-arm trial and not to an adaptive design in which arms may be dropped using pre-specified rules. Such designs also offer greater efficiency whilst minimising the number of participants needed to be randomized. Reporting guidelines for adaptive trials will be covered by the Adaptive designs CONSORT Extension (16).

Sample size

Item 7a

Standard CONSORT item: how sample size was determined

Extension for multi-arm trials: planned sample size with details of how it was determined for each primary comparison

Examples

"Sample size calculations were based on the assumption that 34% of placebo-treated patients and 54–64% of tadalafil-treated patients (once daily and on demand) would achieve an IIEF-EF score

[International Index of Erectile Function-Erectile Function] after DFW [drug-free washout]. A sample size of 412 randomised patients provided 84% power to detect a 20% difference in proportions in the two pairwise comparisons of tadalafil (once daily and on demand) versus placebo (20% drop-out rate assumed)." (17)

"Because a high degree of benefit would be needed to change routine clinical practice, we specified a 3.3% absolute reduction on the basis of estimated incidence in the control group of 11% (30% relative reduction; odds ratio [OR] 0.67). With 90% power and 2.5% significance level to account for the two comparisons, and allowing for an attrition rate of 15%, we needed to recruit 2,345 participants in each group (7,035 participants overall). Two comparisons of equal importance were tested in the trial: silver alloy catheters versus PTFE [polytetrafluoroethylene] catheters and nitrofural catheters versus PTFE catheters." (18)

Explanation

The sample size for a multi-arm trial should correspond to the planned primary comparisons (Item 2b). The approach to sample size is determined by the structure of the interventions being compared and the nature of the planned analyses (Box 1). When pairwise comparisons are planned, the sample size will usually be determined to give adequate power to evaluate each of the intended primary comparisons. If investigators deem that they need to adjust for multiple comparisons, the planned sample size may be inflated to take account of that adjustment (Item 7a).

Statistical methods

Item 12a

Standard CONSORT item: statistical methods used to compare groups for primary and secondary outcomes

Extension for multi-arm trials: explicitly state if no adjustments for multiplicity were applied; if adjustments were made, state the method used.

Examples (where adjustment was not made)

"The hypotheses were that the high-rate group, the delayed-therapy group, or both would have a reduced risk of a first occurrence of inappropriate therapy, as compared with the conventional-therapy group. The two trials were conducted in parallel, with inference made in each, and no adjustment for multiple comparisons was deemed appropriate." (19)

"All p-values are two-sided with no adjustment made for multiple comparisons." (20)

Examples (where adjustment was made)

"We assessed urinary tract infection outcomes with logistic regression and summarised findings as absolute percentage risk differences and ORs, both with 95% CIs calculated as 97·5% confidence intervals to adjust for the two comparisons. For the primary analysis, p=0·025 was regarded as significant." (18)

"For both [Visual Analogue Scale] VAS-immediate pain ratings and pressure data, if the Shapiro-Wilk normality test was passed, repeated measures one-way ANOVA [Analysis of variance] with Bonferroni correction post-hoc pairwise comparisons was conducted to explore any significant difference (P<0.05) between the test conditions." (21)

"… we calculated that we would need to enroll 810 patients in each group for the study to have 90% power to show the superiority of apixaban over placebo, at a two-sided alpha level of 0.05, with the use of the Hochberg multiple-testing method." (22)

Explanation

In general, multi-arm trial analysis strategies may have two broad objectives. First, investigators examine variation in efficacy among several interventions, which can be addressed by an overall

analysis comparing all groups at once. Such an analysis is unlikely to be fully satisfactory, as it will not indicate where the differences lie. Second, two or more specific pairwise comparisons can be made between particular treatments, as described above. In a particular trial, both types of analysis may be performed. Indeed, one strategy (commonly recommended in analysis of agricultural and other experiments) is first to perform a global statistical test across all groups, and only to proceed to paired comparisons if the global test is statistically significant. This strategy does not seem especially desirable for the analysis of clinical trials, where we expect a more focused approach to the evaluation of treatment comparisons.

Two further complications may be present. First, two (or more) of the treatments may be different doses or durations of the same drug or other intervention. In such cases it may be of most interest to examine whether there is a dose response relation rather than simply testing the significance of differences between pairs of treatments. Second, two of the groups may receive variants of the same basic intervention. For example, they may receive the same drug either orally or intravenously. Sensibly, investigators might first compare these groups combined versus the comparison group (usually placebo or standard treatment) before considering whether the two variants might differ. Groups receiving different doses may also sometimes be considered in this way. Where such an analysis is planned it may sometimes be felt that the groups should be allocated in the ratio 1:1:2 to maximise the power of the first comparison.

Statistical adjustment for multiple comparisons invokes debate among methodologists, and there is no consensus. While some would use such an adjustment, others would never apply adjustments (23, 24).

Investigators may avoid multiplicity problems with analytical approaches. Some examples include:

- Using a single global test of significance across comparison groups (e.g., comparing A versus B versus C in a 3-arm trial) and shunning multiple comparisons. Of note, as

mentioned earlier, a single global test across all the treatments, however, is of limited use (25).

- Modelling a dose-response relationship and eliminating multiple comparisons (26).
- Using a prioritized sequence of tests. For example, investigators might decide upon the 300 mg new antibiotic versus standard as the priority test and, if that comparison is statistically significant, only then continue to the 200 mg versus standard comparison. A prioritized sequence of tests addresses multiplicity without adjustments (27).
- Not making adjustment for multiplicity while transparently reporting all comparisons made. Many multi-arm trials are designed for direct comparison of unrelated treatments with a control arm, such as comparing A versus C and B versus C in a three-arm trial. Adjustments for multiple comparisons generally need not play a role in such multi-arm trials (2, 28-30).

Sometimes formal adjustments for multiplicity are unavoidable; some regulators take a firm position. It has been stated that control of the study-wise type I error is a minimal prerequisite for confirmatory claims for drug licensing purposes (31). However, even when adjustment becomes appropriate, implementation becomes problematic. Bonferroni adjustments are often recommended, usually because of their simplicity. However, other adjustment strategies sometimes perform better on the overall control of the type-1-error rate (usually called the family-wise type-1-error (FWER)) (28, 32-34), while performing worse on the probability of more than one false-positive (28). The adjustments frequently provide over-correction for multiplicity, especially using Bonferroni adjustment. It becomes overly conservative as the correlation among the comparisons becomes higher. Other approaches, including Holm, Hochberg, Dunnett's t and the adjusted Hochberg methods, have been compared to the Bonferroni approach (28). All appear less conservative than Bonferroni.

Results

Recruitment

Item 14a

Standard CONSORT item: dates defining the periods of recruitment and follow-up

Extension for multi-arm trials: if different across treatment groups (e.g. groups were added or dropped), periods of recruitment and follow-up and reason(s), and any statistical implications.

Example

Methods (Study Setting and Design): "The study was conducted between January 2001 and December 2004 at 57 clinical sites in the United States (16 university clinics, 10 state mental health agencies, 7 Veterans Affairs medical centers, 6 private nonprofit agencies, 4 private-practice sites, and 14 mixed-system sites). Patients were initially randomly assigned to receive olanzapine, perphenazine, quetiapine, or risperidone under double-blind conditions and followed for up to 18 months or until treatment was discontinued for any reason (phase 1). (Ziprasidone was approved for use by the Food and Drug Administration [FDA] after the study began and was added to the study in January 2002 in the form of an identical-appearing capsule containing 40 mg)

Methods (Statistical Analysis): … Ziprasidone was added to the trial after approximately 40 percent of the patients had been enrolled … and comparisons involving the ziprasidone group were limited to the cohort of patients who underwent randomization after ziprasidone was added (the ziprasidone cohort). In general, the trial had a statistical power of 85 percent to identify an absolute difference of 12 percent in the rates of discontinuation between two atypical agents; however, it had a statistical power … of 58 percent for comparisons involving ziprasidone … The overall difference among the olanzapine, quetiapine, risperidone, and perphenazine groups was evaluated with the use of a test with 3 degrees of freedom (df). If the difference was significant at a P value of less than 0.05, the three atypical-drug groups were compared with each other by means of step-

down or closed testing, with a P value of less than 0.05 considered to indicate statistical significance … The ziprasidone group was directly compared with the other three atypical-drug groups and the perphenazine group within the ziprasidone cohort by means of a Hochberg adjustment for four pairwise comparisons. The smallest resulting P value was compared with a value of 0.013 (0.05÷4). [reiterated in a footnote to Table 2 and Figure 2 legend relating to Outcome Measures of Effectiveness in the Intention-To-Treat (ITT) population]

Results (Discontinuation of Treatment): … Within the cohort of 889 patients who underwent randomization after ziprasidone was added to the trial, those receiving olanzapine had a longer interval before discontinuing treatment for any cause than did those in the ziprasidone group (hazard ratio, 0.76; P=0.028). However, this difference was not significant after adjustment for multiple comparisons (required P value, ≤0.013)." (15)

Explanation

Incorporating an emerging therapy as a new randomization arm in a clinical trial that is open to recruitment would be desirable to researchers, regulators and patients to ensure that the trial remains current, new treatments are evaluated as quickly as possible, and the time and cost for determining optimal therapies is minimised (35). There are numerous methodological and statistical implications that should be considered. These include (i) Family-Wise Error Rate control due to stage effects and multiplicity, (ii) ensuring that only concurrent control group data are used for an unbiased comparison with the added arm(s) (36), (iii) statistical power (comparison with concurrent control group data will require adequate power), (iv) the allocation ratio and/or length of recruitment into each group (improved efficiency could be realised by adjusting the total number of participants required and time spent recruiting to answer the primary hypotheses), (v) potential changes to the control group (it is possible through time, that the existing control group may be shown to be inferior, therefore it is theoretically possible that the control group may have to be changed), and (vi) logistical considerations (e.g. extra funding, the time taken for all necessary

approvals/amendments, sourcing drug, updating trial randomization and clinical database systems, possible impact on blinding, trial oversight, recruitment and so on) (35). The extent to which these need to be considered clearly depends upon the nature and structure of the trial. There is clearly potential for overlap here with the CONSORT extension for adaptive designs (16).

If recruitment into more than one treatment group in a multi-arm trial is stopped prematurely, it is important to give the reasons, since those reasons may differ. In addition, regarding Standard CONSORT item 15: A table showing baseline demographic and clinical characteristics for each group, in the situation where recruitment to all treatment groups is not contemporaneous, one could opt for a single or multiple baseline tables. Again, the important reporting message is that authors must clearly state which participants are included in which comparisons for each group.

Outcomes and estimation

Item 17a

Standard CONSORT item: for each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)

Extension for multi-arm trials: Results for each pre-specified comparison of treatment groups

Example

"Primary Outcomes – At 6 months, the AVVQ [Aberdeen Varicose Veins Questionnaire] score in the foam group was significantly higher (indicating a worse disease-specific quality of life) than that in the surgery group, but the difference was moderate (effect size, −1.74; 95% confidence interval [CI], −2.97 to −0.50; P=0.006). The improvement in the AVVQ score in the laser group did not differ significantly from that in the surgery group. There were no significant differences between the groups in the EQ-5D score [a standardized instrument for measuring generic health status] or the SF-36 [Short Form Health Survey] physical component score. For the post hoc analysis of treatment

with laser versus foam, the only significant difference was in the SF-36 mental component score, which was slightly higher (better generic quality of life) in the laser group than in the foam group (effect size, 1.54; 95% CI, 0.01 to 3.06; P=0.048) … Secondary Outcomes – Quality of Life. At 6 weeks, significant between-group differences (P<0.005) included a lower AVVQ score (indicating a better disease-specific quality of life) in the surgery group than in the foam group (effect size, −2.3; 95% CI, −3.7 to −0.9) and lower SF-36 scores (indicating a worse generic quality of life) in the surgery group than in the laser group for the domains of bodily pain (effect size, −2.7; 95% CI, −4.4 to −0.9), vitality (effect size, −2.3; 95% CI, −3.9 to −0.8), role limitations due to emotional health (effect size, −2.4; 95% CI, −4.0 to −0.8), and role limitations due to physical health (effect size, −3.5; 95% CI, −5.2 to −1.8). These four SF-36 domain scores did not differ significantly (with P<0.005 considered to indicate statistical significance) between groups at 6 months. For the post hoc comparisons of laser treatment versus foam treatment, only the EQ-5D score was significantly lower (indicating a worse generic quality of life) in the foam group at 6 weeks (0.044; 95% CI, 0.014 to 0.074)." (37)

Explanation

Investigators should plan the comparisons intended, document them in the protocol and statistical analysis plan, and report them all in the trial report with appropriate interpretations. If arms have been added or dropped during the trial, it is important that the analysis addresses the implications of doing so. If investigators employed measures to control the overall significance level e.g. if they conducted a single global test of significance across comparison groups, modelled a dose-response relationship or used a prioritized sequence of tests, those details should be reported. If they conducted an analysis that dictated formal adjustments for multiplicity, those methods and limitations should be reported. As discussed previously (item 12a), many multi-arm trials will not employ formal adjustments for multiplicity. In those cases, investigators should still transparently report all comparisons undertaken, planned and unplanned, and provide appropriate interpretations of the results.

Discussion

Limitations

Item 20

Standard CONSORT item: Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses

While no specific extension to the standard CONSORT item is recommended here, authors should address the strengths and limitations of multi-arm trials with regards to issues detailed in Box 1 .

Discussion

Multi-arm trials need careful thought and planning. They offer the opportunity to address more than one research question, can accelerate the discovery of new interventions and facilitate head-to-head comparisons with competing treatment options, potentially resulting in patient benefit whilst optimising use of resources. They can be more appealing to participants and clinicians because typically there is an increased probability of receiving an experimental intervention rather than standard care. However, trialists should always be mindful that the efficiency advantages of multi-arm trials and opportunity to evaluate more interventions quicker, are contingent upon recruiting and collating outcomes on the requisite number of participants.

Multi-arm randomized trials are common and so it is important that reports of these trials include information on features specific to the design to allow readers to make an accurate assessment of the conduct of the trial and interpretation of the results. Transparent and complete reporting is an essential prerequisite for reproducibility. Good reporting also facilitates the inclusion of these trials

in systematic reviews. However, multi-arm trials, especially those with more than three treatment arms, are challenging to design and analyse.

In this Special Communication, we have proposed an extension to the widely adopted CONSORT Statement to enable the full and accurate reporting of multi-arm randomized trials. Such trials require clear objectives and hypotheses referring to all of the treatment arms, and the primary comparisons must be identified in advance. The sample size should be clearly pre-specified and the issue of adjustment for multiple testing should at least be acknowledged. If different across treatment groups (e.g. groups were added or dropped), periods of recruitment and follow-up (and reasons) should be reported, and any statistical implications addressed.

Multiplicity adjustment for multiple comparisons among groups in a multi-arm randomized trial remains a challenging issue. Many multi-arm trials are conducted for efficiency reasons. They compare distinct treatments/interventions against a single control group which could easily have been done in multiple separate trials rather than a single multi-arm trial. Indeed, for a multi-arm trial design where several experimental interventions share a control arm, the trial is focused on evaluating the research question for each intervention separately. The interpretation of the results of one comparison ordinarily has no direct bearing on the interpretation of the others. Many trialists/methodologists argue that multiplicity adjustments are not necessary in such instances, as such adjustments would not be necessary if the treatments/interventions were compared in those separate trials (2, 28-30, 38, 39). Some multi-arm trials evaluate several different doses of the same agent against a control arm. These represent related comparisons. In such situations, trialists and methodologists tend to recommend multiplicity adjustments (2, 28, 29, 34). Indeed, an obvious example would occur with certain decision-making criteria in submissions to a regulatory agency for drug approval. If the sponsor specifies more than one treatment comparison and proposes to claim a treatment effect if one or more of the doses are statistically significant, most trialists and

methodologists suggest an adjustment for multiplicity (2, 28-30, 38, 39). Sweeping declarations of 'always' or 'never' needing to adjust for multiple testing should be ignored – it depends upon the objectives, design and analysis.

Some multi-arm trials may also have other special features, for example, they may be crossover, cluster, or factorial trials. For such trials, the specific recommendations for both types of trial will be relevant. Use of the CONSORT Statement for the reporting of two-group parallel trials has been shown to be associated with improved quality of reporting (40). It is hoped that the routine use of this proposed extension to the CONSORT Statement will result in similar improvements.

The CONSORT Group will continue to monitor and revise its recommendations and is also developing checklists and flow diagrams to help improve the quality of reporting of clinical trials of various designs. Other similar extensions and updates are in preparation, and the most up-to-date versions of all CONSORT recommendations can be found on the CONSORT website (www.consort-statement.org).

Article information

Acknowledgments

Author contributions: Concept and design: all authors. Drafting of the manuscript: all authors. Critical revision of the manuscript for important intellectual content: all authors. Obtaining funding: not applicable. Supervision: all authors. EJ and SH had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflict of interest disclosures: Authors EJ, SH and KS have completed and submitted the ICMJE form for disclosure of potential conflicts of interest (DA died on June 3rd, 2018). DA, SH and KS are authors of the CONSORT 2010 Statement. No other disclosures were reported.

Funding/Support: There was no specific funding for this work; EJ, DA and SH were supported by the University of Oxford, KS was supported by FHI 360 and The University of North Carolina School of Medicine, Chapel Hill, North Carolina.

Role of the funder/sponsor: Not applicable.

Disclaimer: The views expressed in this publication are those of the authors and not necessarily those of the University of Oxford, FHI 360, or The University of North Carolina at Chapel Hill.

Additional contributions: We gratefully thank the members of the CONSORT Group; Diana Elbourne, PhD, Medical Statistics Department, London School of Hygiene & Tropical Medicine; Robert M. Golub, M.D., Associate Professor of Medicine, Northwestern University Feinberg School of Medicine; Trish Groves, Robert Brian Haynes, M.D., PhD, Department of Health Research Methods, Evidence and Impact, McMaster University; John P.A. Ioannidis, M.D., DSc, Meta-Research Innovation Center at Stanford (METRICS), Stanford University; David Moher, PhD, Centre for Journalology, Clinical Epidemiology Program, Ottawa Hospital Research Institute, School of Epidemiology and Public Health, Ottawa; Cynthia Diane Mulrow, M.D., MSc, University of Texas Health Science Center at San Antonio; American College of Physicians; Drummond Rennie, M.D., FRCP, MACP, PR Lee Institute for Health Policy Studies, University of California San Francisco; and especially Matthew R Sydes, MSc,

MRC Clinical Trials Unit at UCL, Institute of Clinical Trials and Methodology, UCL, London, for their helpful comments on an earlier draft of this manuscript. Likewise, we would like to thank Julia Mary Brown, BSc, MSc, Leeds Institute of Clinical Trials Research, University of Leeds and Louise Linsell, BSc, MSc, DPhil, National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford. We would also like to thank Ayodele Odutayo, M.D., DPhil, Applied Health Research Centre, St Michaels Hospital, University of Toronto for providing supplementary information.

We also thank Michael James Bradburn, BSc, MSc, Clinical Trials Research Unit, ScHARR, University of Sheffield, Dena R Howard, BSc, MSc, PhD, Leeds Institute of Clinical Trials Research, University of Leeds and Simon Day, PhD, Director, Clinical Trials Consulting & Training Limited, Buckinghamshire, for their most helpful comments and suggestions. Finally, we thank Andrew Robert King, BA (Hons), and Jenny Shilton Osborne, BSc, MSc, National Perinatal Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, for proof-reading and reproducing the figure respectively. No one was compensated for their contribution.

Box 1: Methodological issues in multi-arm randomized trials

Design	Research Objectives Trials with more than two treatment arms will generally either address a more complex question than a two-arm trial, or more commonly, will attempt to address research questions about more than one intervention. Authors should explicitly define the objectives of a multi-arm trial, referring to all the arms of the study and pre-specifying all planned comparisons of intervention groups, in order to partly mitigate against the effects of multiplicity and accusations of data dredging (i.e. unplanned exploratory analysis). Eligibility Criteria In trials involving multiple drugs, safety/toxicity profiles may reduce the pool of potential participants, adversely affecting recruitment and generalisability. Patient and/or recruiting centre characteristics, or lack of equipoise or resources may preclude randomization to one of the options. A multi-arm trial could include two research treatments with contra-indications but allow patients to be randomized to other arms. For example, not everyone might be suitable for a type of surgery but could contribute to an evaluation of a drug, so patients may be randomized: control versus surgery versus drug, or control versus surgery (if not suitable for drug therapy) or control versus drug (if not suitable for surgery). Sample Size The sample size for a multi-arm phase 3 trial should depend upon the planned primary comparison(s). The sample size per group should be large

	enough so that pre-specified primary comparisons would have adequate power. Use of Placebos for Blinding In the multi-arm design, blinding needs to ensure that none of the arms can be identified. If route of administration of two experimental drugs varies (e.g. oral versus intravenous), blinding can become invasive, expensive and an additional burden on participants. As the number of experimental drug arms increases, blinding may become more problematic. Three-arm trial; placebo and active control Three-arm trials that include an active control group as well as a placebo group can establish whether a failure to distinguish a test treatment effect from placebo implies ineffectiveness of the new test treatment or is simply the result of a trial that lacked the ability to identify an active drug. The comparison of placebo to the active control (standard drug) in such a design provides internal evidence of assay sensitivity (a property of a clinical trial defined as the ability of a trial to distinguish an effective treatment from a less effective or ineffective intervention). If this is considered important, one could employ an unequal allocation ratio in order to make the active groups larger than the placebo group to improve the precision of the active drug comparison. This may increase acceptability to participants and investigators, since there is a lower probability of being allocated to placebo (41).
Conduct	Interim Analysis and Stopping Guidelines Many trials employ formal methods for interim monitoring and 'early' stopping guidelines. These prompt consideration for recruitment to stop early for strong evidence of a benefit, harm or alternatively, futility. Multiple

	treatment arms add to the complexity. Depending on the type/structure of a multi-arm trial, an ethical dilemma may arise as a result of an interim analysis for example, if sufficiently strong evidence of a benefit of one of the treatment interventions versus control is observed. This intervention will be considered a significant improvement over the control arm and recruitment into the control arm may have to be stopped. This may then result in recruitment to the other treatment intervention arms stopping because of a lack of a concurrent control group. Since the trial may be stopped if any of the treatment intervention-control comparisons cross an efficacy 'early' stopping boundary, multiplicity adjustment is required for the efficacy boundaries.
Analysis	Analysis Strategy If the main objective is to examine whether the interventions differ, but not how they differ, it would be appropriate to compare all groups at once using a single global test of significance. If the main objective is to examine a trend, one should model a dose-response. More often, two or more specific comparisons are made between particular pairs or combinations of treatments. However, the number of possible comparisons can be considerable. Multiple Treatments Comparisons For a three-arm trial ('treatments' A, B, and C, say) there are several possible comparisons including: 1. Comparing all three groups at once (A versus B versus C); a global test of unordered groups or a test for trend across ordered groups.

	2. Comparing one group to the other two groups combined (A plus B versus C) and then (A versus B); A and B might be low and high doses of the same drug; the first comparison could be of treated versus untreated, followed by a comparison of the two treated groups; or A and B might be two antibiotics in the same class versus C as a member of a different class (NB. the labelling in this example is arbitrary). 3. All pairwise comparisons: (A versus B), (A versus C) and (B versus C). 4. Comparing (A versus C) and (B versus C), but not (A versus B); for example, comparing two treatments separately to the control, but not comparing the two treatments to each other.
Reporting and Interpretation	Multi-arm trials often address complex and intricate questions concurrently, and as such, have a different focus than two arm trials. For example, following a pre-specified comparison of all groups, the interpretation of a statistically significant global test is not straightforward. The investigators have evidence to reject the hypothesis that all the interventions were equally effective, but no clear indication of precisely where the differences lie. It is tempting, but incorrect, simply to use the observed data to draw more precise conclusions. It is incorrect, for example, to deduce that the intervention with the most favourable results is better than the others, as this question has not been examined explicitly. Moreover, multiple pairwise comparisons may yield apparently paradoxical results. For example, in a trial of two active interventions A and B versus placebo, it is possible to find that A is significantly better than placebo, but

	that B is not significantly different from either A or placebo. It is also possible that no pairwise comparison is significant despite a significant global test. These problems are well known in agricultural and other research areas where formal multi-arm comparisons are common, but there is rather little experience of such issues in clinical research. Lastly, and of general relevance, are interpretation issues relating to the multiplicity of comparisons. Clinicians frequently find the addition of a group to a trial enhances rather than diminishes the information gained(30). In many such trials, interpretation of results adjusted for multiplicity frequently causes rather than solves interpretational problems. Yet, sometimes a particular analysis dictates adjustment for multiple comparisons; if those adjustments are indeed unsophisticated and liable to over-correction, the authors should account for that in their interpretation. Readers of a report of a multi-arm trial will expect a description of how the primary and secondary comparisons were handled, emanating from the multiple intervention groups. Most authors and readers would be likely to bear in mind the number of analyses performed regardless of whether any formal adjustment is made.

Table 1: Checklist for reporting of multi-arm parallel group randomized trials: extension of the CONSORT 2010 statement *

Section/Topic	Item No	CONSORT 2010 Statement checklist item	Multi-arm trial extension
Title and abstract			
	1a	Identification as a randomized trial in the title	Identification as a multi-arm randomized trial in the title or an indication of the number of treatment groups that the participants were randomly assigned to
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts) (42)	Specification of the number of treatment groups; details of any groups added or dropped
Introduction			
Background and objectives	2a	Scientific background and explanation of rationale	Rationale for using a multi-arm design
	2b	Specific objectives or hypotheses	Specification of the research question referring to all of the treatment groups Clear statement of all hypotheses to be tested and primary comparisons involved
Methods			
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	Specification of the number of treatment groups
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	Details of any treatment groups added or dropped (if relevant) with reasons, or changes to allocation ratio
Participants	4a	Eligibility criteria for participants	
	4b	Settings and locations where the data were collected	
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	
	6b	Any changes to trial outcomes after the trial commenced, with reasons	
Sample size	7a	How sample size was determined	Planned sample size with details of how it was determined for each primary comparison
	7b	When applicable, explanation of any interim analyses and stopping guidelines	
Randomization			

Sequence generation	8a	Method used to generate the random allocation sequence	
	8b	Type of randomization; details of any restriction (such as blocking and block size)	
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	
	11b	If relevant, description of the similarity of interventions	
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	Explicitly state if no adjustments for multiplicity were applied; if adjustments were applied, state rationale and the method used
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	
Results			
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	
	13b	For each group, losses and exclusions after randomization, together with reasons	
Recruitment	14a	Dates defining the periods of recruitment and follow-up	If different across treatment groups (e.g. groups were added or dropped), periods of recruitment and follow-up and reason(s), and any statistical implications
	14b	Why the trial ended or was stopped	
Baseline data	15	A table showing baseline demographic and clinical characteristics for each group	
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	Results for each pre-specified comparison of treatment groups
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	

Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms) (43)	
Discussion			
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	
Other information			
Registration	23	Registration number and name of trial registry	
Protocol	24	Where the full trial protocol can be accessed, if available	
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders	

*It is strongly recommended that this checklist is read in conjunction with the CONSORT 2010

Statement Explanation and Elaboration (5) for important clarification on the items.

References

1. Parmar MK, Carpenter J, Sydes MR. More multiarm randomised trials of superiority are needed. Lancet (London, England). 2014;384(9940):283-4.
2. Freidlin B, Korn EL, Gray R, Martin A. Multi-arm clinical trials of new agents: some design considerations. Clinical cancer research : an official journal of the American Association for Cancer Research. 2008;14(14):4368-71.
3. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. Jama. 1994;272(2):122-4.
4. Odutayo A, Emdin CA, Hsiao AJ, Shakir M, Copsey B, Dutton S, et al. Association between trial registration and positive study findings: cross sectional study (Epidemiological Study of Randomized Trials-ESORT). BMJ (Clinical research ed). 2017;356:j917.
5. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ (Clinical research ed). 2010;340:c869.
6. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ (Clinical research ed). 2010;340:c332.
7. Ahren B, Johnson SL, Stewart M, Cirkel DT, Yang F, Perry C, et al. HARMONY 3: 104-week randomized, double-blind, placebo- and active-controlled trial assessing the efficacy and safety of albiglutide compared with placebo, sitagliptin, and glimepiride in patients with type 2 diabetes taking metformin. Diabetes care. 2014;37(8):2141-8.
8. Agar MR, Lawlor PG, Quinn S, Draper B, Caplan GA, Rowett D, et al. Efficacy of Oral Risperidone, Haloperidol, or Placebo for Symptoms of Delirium Among Patients in Palliative Care: A Randomized Clinical Trial. JAMA internal medicine. 2017;177(1):34-42.
9. Connick P, De Angelis F, Parker RA, Plantone D, Doshi A, John N, et al. Multiple Sclerosis-Secondary Progressive Multi-Arm Randomisation Trial (MS-SMART): a multiarm phase IIb randomised, double-blind, placebo-controlled clinical trial comparing the efficacy of three neuroprotective drugs in secondary progressive multiple sclerosis. BMJ open. 2018;8(8):e021944.
10. Dickersin K, Manheimer E, Wieland S, Robinson KA, Lefebvre C, McDonald S. Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials. Evaluation & the health professions. 2002;25(1):38-64.
11. Geddes JR, Goodwin GM, Rendell J, Azorin JM, Cipriani A, Ostacher MJ, et al. Lithium plus valproate combination therapy versus monotherapy for relapse prevention in bipolar I disorder (BALANCE): a randomised open-label trial. Lancet (London, England). 2010;375(9712):385-95.
12. Foa EB, McLean CP, Zang Y, Rosenfield D, Yadin E, Yarvis JS, et al. Effect of Prolonged Exposure Therapy Delivered Over 2 Weeks vs 8 Weeks vs Present-Centered Therapy on PTSD Symptom Severity in Military Personnel: A Randomized Clinical Trial. Jama. 2018;319(4):354-64.
13. Gray R, Ives N, Rick C, Patel S, Gray A, Jenkinson C, et al. Long-term effectiveness of dopamine agonists and monoamine oxidase B inhibitors compared with levodopa as initial treatment for Parkinson's disease (PD MED): a large, open-label, pragmatic randomised trial. Lancet (London, England). 2014;384(9949):1196-205.
14. Howard RJ, Juszczak E, Ballard CG, Bentham P, Brown RG, Bullock R, et al. Donepezil for the treatment of agitation in Alzheimer's disease. The New England journal of medicine. 2007;357(14):1382-92.
15. Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. The New England journal of medicine. 2005;353(12):1209-23.
16. Adaptive designs CONSORT Extension (ACE) Project. Development of a CONSORT Extension for adaptive clinical trials. 2016. Available from: http://www.equator-network.org/wp-content/uploads/2017/12/ACE-Project-Protocol-v2.3.pdf [Accessed 14 Feb 2018].

17. Montorsi F, Brock G, Stolzenburg JU, Mulhall J, Moncada I, Patel HR, et al. Effects of tadalafil treatment on erectile function recovery following bilateral nerve-sparing radical prostatectomy: a randomised placebo-controlled study (REACTT). European urology. 2014;65(3):587-96.
18. Pickard R, Lam T, MacLennan G, Starr K, Kilonzo M, McPherson G, et al. Antimicrobial catheters for reduction of symptomatic urinary tract infection in adults requiring short-term catheterisation in hospital: a multicentre randomised controlled trial. Lancet (London, England). 2012;380(9857):1927-35.
19. Moss AJ, Schuger C, Beck CA, Brown MW, Cannom DS, Daubert JP, et al. Reduction in inappropriate therapy and mortality through ICD programming. The New England journal of medicine. 2012;367(24):2275-83.
20. Ndibazza J, Mpairwe H, Webb EL, Mawa PA, Nampijja M, Muhangi L, et al. Impact of anthelminthic treatment in pregnancy and childhood on immunisations, infections and eczema in childhood: a randomised controlled trial. PloS one. 2012;7(12):e50325.
21. Fong DT, Pang KY, Chung MM, Hung AS, Chan KM. Evaluation of combined prescription of rocker sole shoes and custom-made foot orthoses for the treatment of plantar fasciitis. Clinical biomechanics (Bristol, Avon). 2012;27(10):1072-7.
22. Agnelli G, Buller HR, Cohen A, Curto M, Gallus AS, Johnson M, et al. Apixaban for extended treatment of venous thromboembolism. The New England journal of medicine. 2013;368(8):699-708.
23. Perneger TV. What's wrong with Bonferroni adjustments. BMJ (Clinical research ed). 1998;316(7139):1236-8.
24. Rothman KJ. No adjustments are needed for multiple comparisons. Epidemiology (Cambridge, Mass). 1990;1(1):43-6.
25. Pocock SJ. Clinical Trials: A Practical Approach. Chichester, UK: John Wiley & Sons Ltd; 1983.
26. Senn S. Statistical Issues in Drug Development. Chichester, UK: John Wiley & Sons Ltd; 1997.
27. Bauer P, Chi G, Geller N, Gould AL, Jordan D, Mohanty S, et al. Industry, government, and academic panel discussion on multiple comparisons in a "real" phase three clinical trial. Journal of biopharmaceutical statistics. 2003;13(4):691-701.
28. Howard DR, Brown JM, Todd S, Gregory WM. Recommendations on multiple testing adjustment in multi-arm trials with a shared control group. Statistical methods in medical research. 2018;27(5):1513-30.
29. Wason JM, Stecher L, Mander AP. Correcting for multiple-testing in multi-arm trials: is it necessary and is it done? Trials. 2014;15:364.
30. Schulz KF, Grimes DA. Multiplicity in randomised trials I: endpoints and treatments. Lancet (London, England). 2005;365(9470):1591-5.
31. European Medicines Agency. Guideline on multiplicity issues in clinical trials [draft]. 2017. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2017/03/WC500224998.pdf [Accessed 14 Feb 2018].
32. Sankoh AJ, D'Agostino RB, Sr., Huque MF. Efficacy endpoint selection and multiplicity adjustment methods in clinical trials with inherent multiple endpoint issues. Statistics in medicine. 2003;22(20):3133-50.
33. Hsu JC. Theroy and Methods. New York: Chapman & Hall; 1996.
34. Proschan MA, Waclawiw MA. Practical guidelines for multiplicity adjustment in clinical trials. Controlled clinical trials. 2000;21(6):527-39.
35. Cohen DR, Todd S, Gregory WM, Brown JM. Adding a treatment arm to an ongoing clinical trial: a review of methodology and practice. Trials. 2015;16:179.
36. Altman DG. Avoiding bias in trials in which allocation ratio is varied. Journal of the Royal Society of Medicine. 2018;111(4):143-4.

37. Brittenden J, Cotton SC, Elders A, Ramsay CR, Norrie J, Burr J, et al. A randomized trial comparing treatments for varicose veins. The New England journal of medicine. 2014;371(13):1218-27.
38. Duncan DB. Multiple range and multiple F tests. Biometrics. 1955;11:1-42.
39. Cook RJF, V. T. Multiplicity considerations in the design and analysis of clinical trials. J R Stat Soc. 1996;159:93-110.
40. Turner L, Shamseer L, Altman DG, Weeks L, Peters J, Kober T, et al. Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. The Cochrane database of systematic reviews. 2012;11:Mr000030.
41. European Medicines Agency. ICH Topic E 10 Choice of Control Group in Clinical Trials. [Internet]. London: European Medicines Agency. 2001. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002925.pdf [Accessed 14 Feb 2018]
42. Hopewell S, Clarke M, Moher D, Wager E, Middleton P, Altman DG, et al. CONSORT for reporting randomized controlled trials in journal and conference abstracts: explanation and elaboration. PLoS medicine. 2008;5(1):e20.
43. Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Annals of internal medicine. 2004;141(10):781-8.