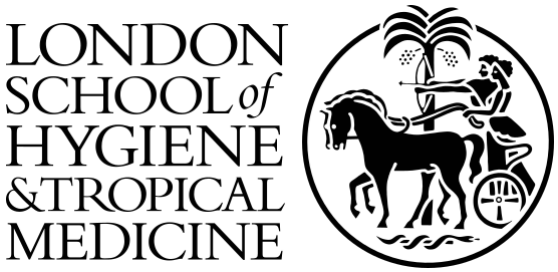


CONSORT 2010 statement: extension to randomised crossover trials

Kerry Dwan, Statistical Editor[1], Tianjing Li, Associate Professor[2], Douglas G Altman, Professor of statistics in medicine[3], Diana Elbourne, Professor of healthcare evaluation[4]

[1] *Review Production and Quality Unit, Editorial and Methods Department, Cochrane Central Executive, Cochrane, St Alban's House, 57-59 Haymarket, London SW1Y 4QX, UK*

[2]*Center for Clinical Trials and Evidence Synthesis, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Baltimore, MD, 21205 USA*

[3]*Centre for Statistics in Medicine, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK*

[4]*London School of Hygiene and Tropical Medicine, Department of Medical Statistics, London WC1E 7HT, UK*

Correspondence to kdwan@cochrane.org

Contributors: DGA and DE initiated the work. KD, TL, DE and DGA drafted the manuscript and all authors reviewed it and KD, TL and DE approved the final version (DGA died in June 2018). KD is the guarantor. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Dr. Dwan and Dr. Elbourne have nothing to disclose. Dr. Li reports grants from National Eye Institute, National Institutes of Health, grants from National Library of Medicine, National Institutes of Health, during the conduct of the study.

Contribution into other languages, create adaptations, reprints, include within collections and create summaries, extracts and/or, abstracts of the Contribution, iii) create any other derivative work(s) based on the Contribution, iv) to exploit all subsidiary rights in the Contribution, v) the inclusion of electronic links from the Contribution to third party material where-ever it may be located; and, vi) license any third party to do any or all of the above."

Abstract

Evidence shows the quality of reporting of randomised controlled trials (RCTs) is not optimal. The lack of transparent reporting impedes readers from judging the reliability and validity of trial findings, prevents researchers from extracting information for systematic reviews, and results in research waste. The Consolidated Standards of Reporting Trials (CONSORT) Statement was developed to improve the reporting of RCTs. The primary focus was on parallel group trials with two treatment groups. Crossover trials are a particular type of trial for chronic conditions in which participants are randomised to a sequence of interventions. They are a useful and efficient design because participants act as their own control. The reporting of crossover trials has, however, been variable and incomplete, hindering their usefulness in clinical decision making and by future researchers. We present the CONSORT extension to randomised crossover trials. It aims to facilitate better reporting of crossover trials. The CONSORT 2010 checklist is revised for crossover designs, and introduces a modified flowchart and baseline table to enhance transparency. Examples of good reporting and evidence-based rationale for CONSORT crossover checklist items are provided.

Introduction

Inadequate reporting of randomised controlled trials (RCTs) is associated with bias in the estimation of treatment effects [1, 2]. It also impairs the critical appraisal of the quality of randomised trials, which is important to assess the validity of the results of the individual trial and in conducting systematic reviews. To attempt to address this issue, the Consolidated Standards of Reporting Trials (CONSORT) statement is a set of recommendations for the reporting of RCTs [3]. It comprises a checklist of essential items that should be included in reports of RCTs and a diagram to document the flow of participants through the trial from before group assignment through to the final analysis. These items are evidence-based whenever possible. Explanation and elaboration of the rationale for checklist items is provided in an accompanying article [4]. Many journals now require that reports of RCTs conform to the recommendations in the CONSORT Statement [5].

The primary focus of the CONSORT Statement is the most common type of RCT with two treatment groups (two 'arms') using an individually randomised, parallel group, superiority design [3]. Almost all the elements of the CONSORT Statement apply equally to RCTs with other designs, but some elements need adaptation, and in some cases additional issues need to be discussed. Members of the CONSORT group have published several extension papers that augment the CONSORT Statement in relation to types of interventions and data. Extensions of CONSORT 2010 to different trial designs have also been published including cluster randomised trials [6], non-inferiority and equivalence trials [7], N-of-1 trials [8], pragmatic trials [9] and within-person trials [10]. As part of that series, in this paper we extend the CONSORT 2010 recommendations to simple crossover RCTs in which participants receive two treatments sequentially over two periods and the order in which treatments are taken is randomised.

Scope of this paper

First, we summarise the key methodological features of crossover trials. Second, we consider the empirical evidence about how common crossover trials are and review published studies of the quality of reporting of such trials. Following these literature reviews, we make suggestions for amendments to the CONSORT checklist adapted for crossover trials and give illustrative examples of good reporting. In this guideline we focus on the simplest and most common form of the randomised crossover trial in which all

participants receive two interventions in one of two sequences (known as the 2x2 or AB/BA design). Most of the recommendations apply also to the more complicated designs (more than two interventions, periods or sequences). Specific issues that arise in trials comparing more than two interventions are briefly discussed in a separate section below.

Methodological features of randomised crossover trials

In contrast to a parallel group trial, each individual in a crossover trial receives multiple interventions but in a random order, that is participants are randomised to sequences of interventions. In this way, each participant acts as his or her own control. Such pre-specified designs should not be confused with trials in which some individuals "cross over" through non-compliance or use of rescue medication, or in which all participants in the control group are given the chance to "cross over" to the experimental treatment at the end of the main trial. Zeng et al found that almost one quarter of records (n=17/72) labelled as 'Crossover Assignment' did not use a randomised crossover design to randomise participants to a sequence; instead, these trials allowed participants to change intervention during the course of the trial [11].

Randomised crossover trials present particular challenges. One challenge is the potential for a 'carry-over effect', that is, the effect of the first intervention persists into the second period such that the observed difference between the treatments depends upon the order in which they were received. A carry-over effect could have a range of causes. As well as the obvious problem of a drug or other treatment remaining in the system, participants' later responses can be affected by previous side effects or other reactions to previous treatment. It is recommended that crossover trials should include a sufficient 'washout' between the end of the first intervention and the start of the second intervention, so that any effects from the first intervention will not be 'carried over' to the measurement of outcome in the second intervention period.

Another issue is the 'period effect' which occurs when the outcome of interest changes with time irrespective of treatment effect, for example, the condition may not be stable or the effect of treatment is seasonal.

A further issue is the possibility of participants dropping out of the trial if the first intervention is either very successful or unsuccessful; the results for these participants cannot be included in the analysis.

Box 1: Glossary

Period: A length of time where one treatment was received

Sequence: treatment sequence (AB, BA), subjects allocated to the AB study arm receive treatment A first, followed by treatment B, and vice versa in the BA arm

Within-participant variability: the expected standard deviation of the within-participant differences

Washout: A length of time between treatment periods where no treatment is received in order to allow the treatment to wear off

Carry-over effect: when the effect of the first intervention persists into the second period.

Period effect: the outcome of interest changes with time irrespective of treatment effect

Within-participant comparison: A within participant comparison takes into account the correlation between measurements for each participant as they act as their own control, therefore measurements are not independent.

Design

The particular strength of the simple AB/BA crossover design is that both interventions are evaluated on the same participant, allowing comparison at the individual rather than group level. In addition, participants in a crossover trial can express preferences by comparing their experiences of the two interventions, which is not possible in a parallel group design as they will only have received one intervention [12].

A crucial methodological question is whether the use of the crossover design is justified. Crossover trials are most appropriate for symptomatic treatment (i.e., treatment for symptoms such as, for example, pain) of conditions or diseases that are chronic or relatively stable (such as multiple sclerosis or rheumatoid arthritis), at least over the time period under study, and when the treatment effects are reversible, and short-lived. The crossover design is inappropriate when the condition of interest can be cured or when participants are likely to die over the period of the trial. The design is quite commonly

used, however, in less appropriate circumstances. For example, pregnancy is an intended outcome of sub-fertility treatment. If a woman becomes pregnant in the first period of the trial (i.e. before crossover), she will be precluded from entry into subsequent phases of the trial. Nevertheless the design is defended in the field [13] (for instance it has been suggested that the pregnancies can be treated statistically as 'missing at random' (see [14])), and remains common despite criticism [15].

The sample size calculation for such trials is based on the within-participant variability in responses. The crossover design is much more efficient than the parallel design when there is a high positive correlation between participants' responses to the different treatments. Compared to a parallel group design, fewer participants are required for a crossover trial to obtain the same power for a target effect size and type 1 error rate.

Crossover trials have certain weaknesses. In particular, there can be carry-over effects as discussed above. Participants may drop out after the first treatment, and so not receive the second treatment. Withdrawal may be related to side-effects.

Analysis

The analysis of a cross-over trial should be based on paired data [16-18]. The estimation approaches should account for the correlation of repeated measurements in the same individual. The tests for significance should utilize procedures such as the paired t test (assuming no carryover/period effect) which is based on within-participant differences for a continuous response and the Mainland-Gart test for a binary response [19, 20].

A previously recommended but criticised method of analysing crossover trials was to test for carryover and if this was significant to discard the second period data and analyse only the data from the first period. In other words, the first period's data are analysed as if from a parallel group trial. Freeman [21] showed that this strategy is seriously flawed, and leads to biased answers (as is generally the case when the choice between two analyses is based on the result of a preliminary hypothesis test). Senn [17] and others have argued that the use of the two period two treatment crossover design is effectively built on the assumption that there is minimal carryover effect.

The other statistical issue specific to crossover studies is the need for adjustment for possible period and carryover effects. Parameters can be included for carry-over effect in

the statistical model. In the AB/BA crossover design, the terms 'carry-over' and 'treatment by period' interaction sometime are used interchangeably because the effects of 'carry-over' and 'treatment by period interaction' are not separately identifiable in the data. Although carry-over effect can be estimated, Senn [17] and others have argued that there is little value of using the carry-over effect to adjust the treatment effect. This is because such adjustment relies on assumptions about the nature of the possible carry-over effect and reduces the statistical efficiency for estimating the main treatment effect.

On the other hand, period effect can be dealt with and adjusted for in the analysis. In the AB/BA crossover design, when equal numbers of participants are allocated to each sequence, then on average the period effect will not bias the estimate of treatment effect. However, a period effect will affect the variance estimate because it interferes with how much of the treatment effect might be attributed to random variation. It is important to present data for readers to understand the extent of period effect and communicate clearly whether the period effect was adjusted for or not in the analysis, and whether such a decision was made *a priori*.

How common are randomised crossover trials?

A detailed review of all PubMed-indexed RCTs published in December 2000 found that 74% (383/519) trials used a parallel design and 22% (116/519) were crossover trials [22] and of the trials indexed in MEDLINE in December 2000, 22% (116/526) were crossover trials and most used two treatments (72%) and had two periods (64%) [23]. A review of all PubMed-indexed RCTs published in December 2006 found 77% (477/616) trials used a parallel design and 16% (100/616) were crossover trials [24]. A review of ClinicalTrials.gov of intervention studies registered between 2007 and 2010 found that 11.2% (4351/38969) were of crossover design [25]. A more recent review of PubMed, in December 2012, found that 8.7% (98/1122) of RCTs were crossover design [26]

What is the quality of reporting of randomised crossover trials?

Although articles on the quality of reporting of RCTs in relation to CONSORT are relatively common, few articles have specifically examined the quality of reporting of crossover

trials. Mills et al. found that randomised crossover trials indexed in MEDLINE in December 2000 frequently omitted details on design, analysis and interpretation [23]. However, most trials reported and defended a washout period (70%, 87/127) and reported use of paired data in the analysis (95%, 121/127). Gewandter et al. investigated 124 crossover clinical trials of pharmacologic treatments for chronic pain published between 1993 and 2013 and found that 28% (35/124) of trials reported baseline and post washout pain levels and only 32% (23/75) reported a sample size calculation that specifically indicated that it was based on within participant variability [27]. Straube et al. considered 98 crossover trials on chronic painful conditions published between 1990 and 2014 and indexed on PubMed and found that adverse events were poorly reported in the abstracts of the trial reports and also infrequently reported in the full article and only 23% (23/98) presented a breakdown by treatment period [28]. Zeng et al [29] found that of 54 phase 3 randomised crossover trials analysed from ClinicalTrials.gov in September 2014, nearly two-thirds were a simple AB|BA design with most trials (87%, 47/54) providing sufficient information for the participant flow throughout the trial. Baseline characteristics were most often reported for all participants as a single group (59%, 32/54) and primary outcomes and adverse events were most commonly reported 'per intervention' (81%, 44/54 and 83%, 45/54 respectively). The reporting of results in Baseline Characteristics, Outcome Measures, and Adverse Events generally did not appear to fully reflect the crossover design.

Several studies have considered the reporting of randomised crossover trials in relation to meta-analyses [30-32] and found that data were frequently reported inappropriately to allow them to be included in a meta-analysis.

These studies show that the problems have not improved over many years and the majority of these studies call for guidance on reporting of randomised crossover trials.

Methods used to develop this CONSORT extension

In May 2002, a number of CONSORT authors met in Arlington, Virginia, USA to consider extensions to the 2001 CONSORT Statement in a range of different designs. The first drafts of a paper extending the Statement to crossover trials were developed by Doug Altman (DA) and Diana Elbourne in 2002-3. In 2010, the CONSORT Statement was updated. Work on the extension to crossover trials progressed in 2014 when Kerry Dwan

and then Tianjing Li joined the group. The checklist and explanatory text were informed by reviews of published randomised trials (as cited above) and completed through numerous teleconferences between the authors from 2014 to 2018. We followed guidance of the CONSORT group to include a member of CONSORT Group Executive (DA), who was also chair of the EQUATOR Steering Group. A draft paper was distributed to the wider CONSORT Group and other selected individuals, and the paper was revised to take account of their feedback, and approved by the Executive.

CONSORT checklist for randomised crossover RCTs

Table 1 shows the suggested modifications to the Standard CONSORT checklist for randomised crossover trials. In this section we discuss the checklist items, focussing on those where there are changes to the standard CONSORT items, explain the background and provide one or more examples of good reporting. We also discuss some other checklist items for which we do not suggest any modification but for which implementation requires specific considerations for crossover RCTs.

Title and Abstract

Item 1a: Title

Identification as a randomised crossover trial in the title.

Standard CONSORT item: Identification as a randomised trial in the title.

Examples

Example 1

"Effect of Ginkgo Biloba on Visual Field and Contrast Sensitivity in Chinese Patients With Normal Tension Glaucoma: A Randomized, Crossover Clinical Trial" [33].

Example 2

"Effects of Unfermented and Fermented Whole Grain Rye Crisp Breads Served as Part of a Standardized Breakfast, on Appetite and Postprandial Glucose and Insulin Responses: A Randomized Cross-over Trial" [34].

Explanation

The primary reason for identifying the design in the title is to help readers to identify the study design. Identification of the trial as a randomised crossover trial also ensures that readers will start thinking of the implications of the design in relation to sample size and analysis.

Item 1b: Abstract

Specify a crossover design and report all information outlined in table 2.

Standard CONSORT item: Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts [3]).

Examples

"CONTEXT:

The relationship between sildenafil citrate use and reported adverse cardiovascular events in men with coronary artery disease (CAD) is unclear.

OBJECTIVE:

To evaluate the cardiovascular effects of sildenafil during exercise in men with CAD.

DESIGN, SETTING, AND SUBJECTS:

Randomized, double-blind, placebo-controlled **two period** crossover trial conducted March to October 2000 at a US ambulatory-care referral center among 105 men **(55 to receive sildenafil first, and 55 to receive placebo first)** with a mean (SD) age of 66 (9) years who had erectile dysfunction and known or highly suspected CAD.

INTERVENTIONS:

All patients underwent 2 symptom-limited supine bicycle echocardiograms separated by an interval of 1 to 3 days after receiving a single dose of sildenafil (50 or 100 mg) or placebo 1 hour before each exercise test.

MAIN OUTCOME MEASURES:

Hemodynamic effects of sildenafil during exercise (onset, extent, and severity of ischemia) assessed by exercise echocardiography.

RESULTS:

The difference between mean change after sildenafil and placebo use was 4.3 (95% CI, 0.9-7.7; P =.01). Exercise capacity was similar with sildenafil use and placebo use (mean difference, 0.07; 95% CI, -0.06 to 0.19; P =.29). Exercise blood pressure and heart rate increments were similar. Dyspnea or angina developed in 69 patients who took sildenafil and 70 patients who took placebo (P =0.89); exercise electrocardiography was positive in 12 patients (11%) who took sildenafil and 17 patients (16%) who took placebo (P =0.09). Exercise-induced wall motion abnormalities developed in similar numbers of patients after sildenafil and placebo use (84 and 86 patients, respectively; P =0.53). Wall motion score index at peak exercise was similar after sildenafil and placebo use (mean difference, 0.01; 95% CI, -0.01 to 0.03; P =.40).

CONCLUSION:

In men with stable CAD, sildenafil had no effect on symptoms, exercise duration, or presence or extent of exercise-induced ischemia, as assessed by exercise echocardiography." [**Adapted from** [35]]

Explanation

Clear, transparent, and sufficiently detailed abstracts are important. Readers may only have access to the abstract, and many others will skim it before deciding whether to read further. A well-written abstract also helps retrieval of relevant reports from electronic databases. In 2008 a CONSORT extension on reporting abstracts of randomised trials was published [36] and those recommendations were incorporated into CONSORT 2010 [3].

Abstracts for crossover RCTs should indicate the design of the trial and therefore the randomisation to sequence, and analysis taking account of the within-participant comparisons. Table 2 shows information to be included in the abstract of a crossover trial.

We were not able to find examples of good reporting tackling all the items required. We have therefore adapted a published abstract (see example).

Methods

Item 3a: Trial design

Rationale for a crossover design. Description of the design features including allocation ratio, especially the number and duration of periods, duration of washout period and consideration of carryover effect.

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

Examples

Example 1

"The trial was a randomised double-blind, placebo controlled, crossover design of 15 months' duration. …. randomisation (1 month); treatment period one (6 months); washout (2 months); and finally treatment period two (6 months)… Patients were randomly assigned azithromycin in treatment period one, followed by placebo in treatment period two, or placebo in treatment period one followed by azithromycin in treatment period two." [37]

Example 2

"A crossover design was chosen for this study instead of the more traditional randomized, parallel-group design because the within-patient variation is less than the between-patient variation and thus required fewer patients. In addition, some of the known disadvantages of the crossover design (e.g., larger dropout rate, instability of the patient's condition, and a potential carryover effect) were not expected in this study." [38]

Example 3

"Each treatment period was separated by a 2-week washout, equating to five or more half-lives for either treatment, to allow the effective systemic elimination of the drug before initiation of subsequent treatment." [39]

Example 4

"We did not include a medicine-free period between treatments to increase patient safety. In addition, we believed the 8-week treatment period was sufficient to allow for

the washout of the first treatment before the efficacy measurements at the end of period 2.” [40]

Explanation

The methods should contain a rationale for the use of a crossover design in the given setting. In particular, given that a carry-over effect can neither be identified with sufficient power, nor can adjustment be made for such an effect in the 2x2 crossover design, the assumption needs to be made that any carry-over effects are negligible and some justification presented for this. The description of the design should make clear how many interventions were tested, through how many periods, including information on the length of the treatment, run-in and washout periods (if any).

Item 3b: Changes to methods

Important changes to methods after trial commencement (such as eligibility criteria), with reasons

No change from standard CONSORT item

Explanation

A test for carryover is not recommended. However, if a test for carryover is performed as a result of which the authors use only the first period data, then this should be reported. The use of the test should also be discussed under item 12a (Statistical methods). The reason explaining the presence of a carry-over should also be discussed.

Item 5: Interventions

The interventions with sufficient details to allow replication, including how and when they were actually administered

Standard CONSORT item: The interventions for each group with sufficient details to allow replication, including how and when they were actually administered

Explanation

For this item, 'for each group' was deleted for the extension as in a 2x2 randomised crossover trial, the intention is that all participants receive both of the interventions.

Item 7a: Sample size

How sample size was determined, accounting for within participant variability

Standard CONSORT item: How sample size was determined

Examples

Example 1

"Earlier research of the Cambridge study site (unpublished data) with the Apathy Evaluation Scale showed a mean score of 31 points (standard deviation SD=15.6). If we define a clinical significant improvement on the AES-I as a 35% reduction of the mean score, this leads to an absolute effect size of 0.35*31 points=10.85 points. Thus a conservative estimate of 10 units is used for sample size estimation. Furthermore a within subjects SD=15.0 is assumed. When the sample size in each sequence group is 19, (a total sample size of 38) a 2 x 2 crossover design will have 80% power to detect a difference in means of 10.000 (the difference between a Treatment 1 mean, µ1, of 31 and a Treatment 2 mean, µ2, of 21) assuming that the crossover ANOVA √MSE is 15.000 (the Standard deviation of differences, sd, is 21.213) using a two group t-test (Crossover ANOVA) with a 0.050 two-sided significance level. In order to account for potential drop-outs 40 patients will be randomized. Sample size calculation was performed with nQuery 7.0" ***[41]***

Explanation

A key advantage of the crossover design is that, for a given significance level, power, and effect size, a smaller sample size is required compared to a parallel design in which each participant receives only one treatment. This is because each participant acts as his/her own control (each participant receives both the experimental and control intervention), so the within-participant variability is removed.

It is important that trial authors report the usual quantities required for sample size calculation, including significance level and power, but also for continuous variables the within-participant variability as shown in Example 1. It is often difficult to get the

necessary within-participant information to inform the sample size calculation. Published reports of crossover trials should clarify how the sample size was determined, and ideally should indicate that an appropriate estimate of within-participant variability was used. For crossover trials with a continuous outcome, it is the expected standard deviation of the within-participant differences that must be incorporated into the sample size estimation. In practice, for many trials it is unlikely that there will be data to support a realistic estimate of this value, yet ignoring it is likely to result in an overestimation of the sample size for a crossover trial and is thus conservative [42]. Some attempt should be made to estimate the standard deviation of the within-participant differences (or allow for the correlation).

Likewise, with a *binary outcome*, not considering the paired nature of the data will result in an unnecessarily large sample size due to failure to account for the within participant comparison arising from the paired design. Authors are expected give appropriate details so that the sample size calculation can be replicated.

Any allowance in the sample calculation for losses to follow-up should also be reported.

Item 8a: Sequence generation

Method used to generate the random allocation sequence.

No change from standard CONSORT item

Examples

Example 1

"After a 4-week placebo run-in, eligible patients were randomly assigned, according to a computer generated allocation schedule, to 1 of 2 treatment sequences: montelukast and placebo-matching salmeterol or salmeterol and placebo-matching montelukast. After a 2-week washout, patients crossed over to the other treatment." [43]

Example 2

"Eligible subjects were randomized in a 1:1 allocation to one of two treatment sequences—denosumab/alendronate or alendronate/denosumab—and received each treatment for 1 year." [44]

Explanation

In crossover RCTs, allocation sequence refers to the order in which interventions are received. The allocation may be to sequence one, in which participants have A followed by B, or to sequence two, in which participants have B followed by A.

Item 10: Implementation

Who generated the random allocation sequence, who enrolled participants, and who assigned participants to the sequence of interventions

Standard CONSORT item: Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions

Explanation

For this item, 'the sequence of' was included before interventions as participants are randomised to a sequence of interventions rather than one intervention.

Item 12a: Statistical methods

Statistical methods used to compare groups for primary and secondary outcomes which are appropriate for crossover design (i.e. based on within participant comparison)

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

Examples

Example 1

"Cross-over analyses for health related quality of life scores averaged the between-treatment difference for each patient within each sequence and then across both sequences, providing an estimate of treatment effect. The estimated treatment difference, 95% CI and P value were adjusted for period and sequence effects in the **analysis of variance model**." [39]

Example 2

"**A generalized linear mixed-models** approach was used to estimate differences between periods of electrical stimulation and no stimulation while accounting for within-subject correlations arising from the crossover design." [45]

Example 3

"Statistical analysis allowed for the comparison of both treatment groups with respect to baseline information and subsequent comparison at 2 and 4 weeks for treatment effect. The investigator's assessment and patient's assessment of treatment were analysed using **Gart's test** for binary responses, which takes treatment order [strictly period] into account." [46]

Example 4

"Side effects and patient preferences were analyzed descriptively and using **McNemar's test**." [47]

Example 5

"**Prescott's test** was used to analyze the primary end point to test the significance of difference between the two treatments in the presence of period effects." [39]

Explanation

In line with recommendations made by the International Committee for Medical Journal Editors (ICMJE) and the CONSORT group, analytical methods should be described "with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results."(http://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html#d, accessed March 2019) Identification of the crossover design and the statistical methods used allows readers to evaluate the methods of analysis.

The analysis of a crossover trial should respect the within-participant nature of the comparisons. The Methods section should specify which method of analysis was used. This should make clear how the within-participant analysis has been constructed, for example using t tests on within-participant differences, or ANOVA with participant, period and treatment effects. If period effects and carryover have been modeled then this should be reported. Likewise, for a binary outcome, conditional logistic regression provides an

alternative way of conducting the Mainland-Gart test. The consequences of an analysis not accounting for a within-participant comparison may overestimate the variance for the treatment effect.

In some crossover trials participants are measured on the outcome variable at the beginning as well as at the end of both periods, and the treatment effect is estimated using the change score from each period. This seemingly intuitive approach is claimed to eliminate carryover effect; however it could produce a less precise and even biased estimate of treatment effect [48, 49] and therefore should be discouraged.

While missing data raise the same generic issues in crossover trials as in other designs, the specifics are more complicated. The analysis model, in the absence of missing data, should be identified and the role of baseline data needs to be carefully considered, since often baseline adjustment increases the standard error. A mixed model of all available data (e.g. in this context, with a mixture of fixed and random effects) is typically the preferred first step, with the contextually appropriate adjustment for within-subject dependence, and is valid under Rubin's 'Missing at Random' assumption. Broadly, this states that the distribution of later outcome data, given treatment sequence and earlier data, is the same whether or not those data are observed. Analysis of the complete records gives a valid ITT estimate assuming the distribution of the outcomes given baseline and treatment sequence is the same, whether or not they are observed (i.e. missing at random). One can explore the robustness of the conclusion to this untestable assumption by multiply imputing the data and forcing the distribution of imputed outcomes to differ from the observed ones given baseline and treatment sequence. The use of multiple imputation, imputing from subsets of patients (rather than single mean imputation, last value carried forward, or best/worst imputation) is welcome because the imputed data is both contextually plausible and appropriately reflects the variability [50].

Results

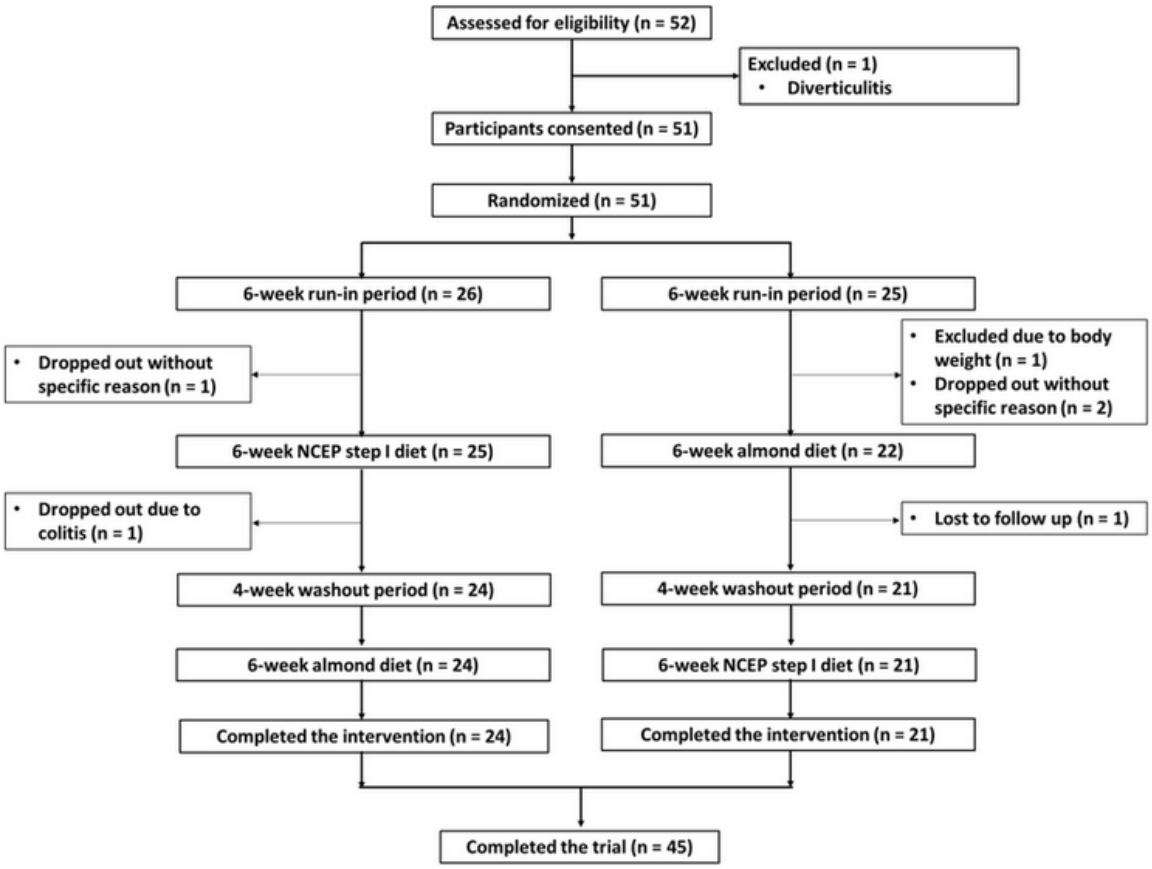
Item 13a: Participant flow (A flow diagram is strongly recommended)

The numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome, separately for each sequence and period. [See Figure 1]

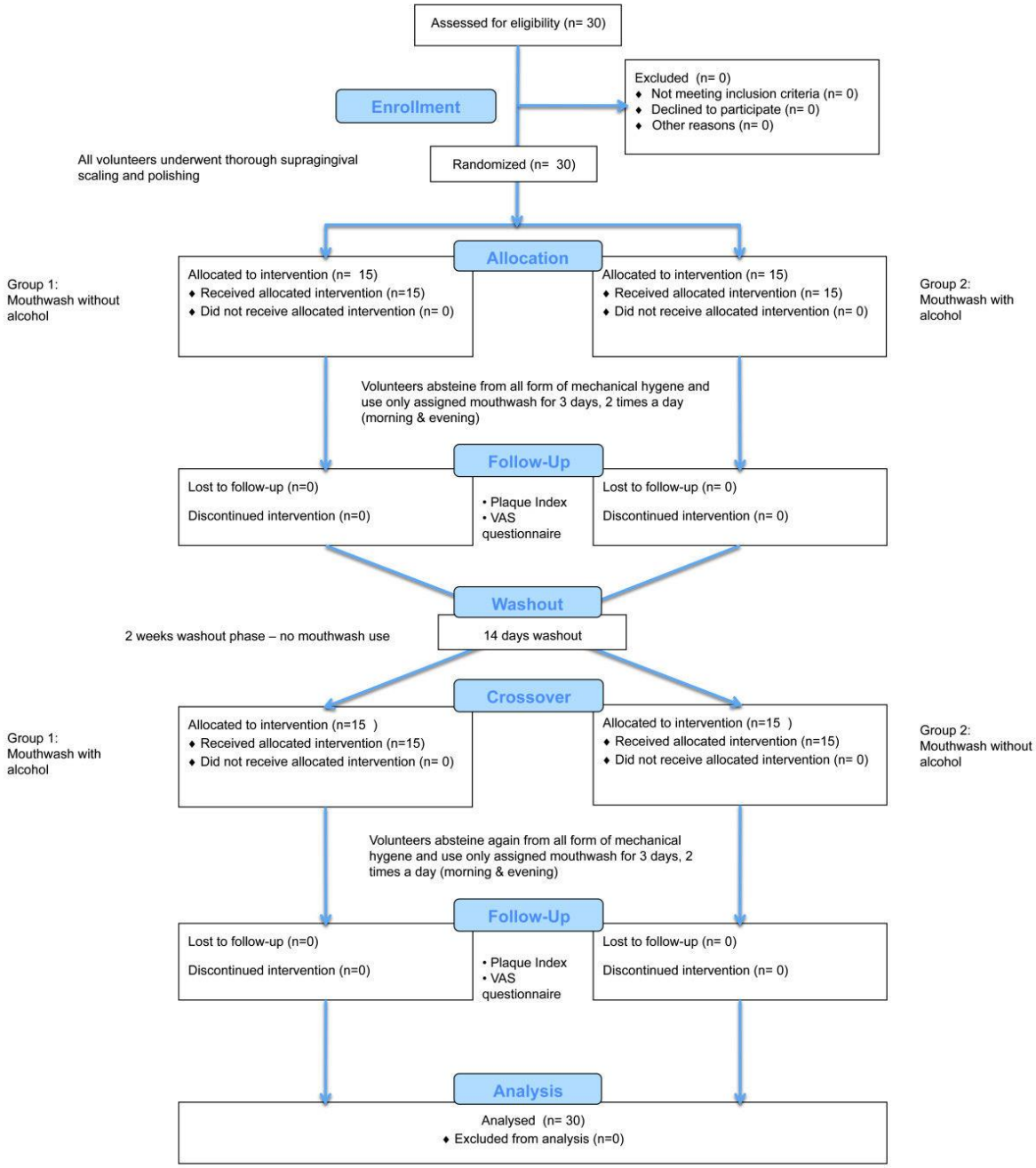
Standard CONSORT item: For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome

Examples

Example 1: [51]



553 Example 2: [52]



554

555 ***Explanation***

556 The flow diagram is a key element of the CONSORT Statement and has been widely
557 adopted. For crossover trials it is important to understand the flow of participants across
558 periods. Although we recommend a flow diagram for communicating the flow of

participants throughout the study, the exact form and content can vary in relation to the specific features of a trial. We recommend using vertical alignment and including a timescale.

Item 13b: Losses and exclusions

Number of participants excluded at each stage, with reasons, separately for each sequence and period.

Standard CONSORT item: For each group, losses and exclusions after randomisation, together with reasons

Examples

Example 1

"One subject assigned to receive active placebo first withdrew because of a scheduling conflict before taking any study medication. Two subjects assigned to receive pregabalin first withdrew in the first period because of adverse events. The remaining 26 subjects completed the study." [53]

Example 2

"Of the 23 patients who provided consent, 17 were randomized to a treatment sequence (9 to pancrelipase then placebo, 8 to placebo then pancrelipase). Sixteen patients completed the study; 1 patient (pancrelipase/ placebo sequence) withdrew consent on day 2 of the first treatment period." [54]

Explanation

A participant who drops out part way through the trial will have their outcome assessed for only one intervention. Dropping out maybe informative; for example, they may be dissatisfied with treatment they were given so do not wish to try any other treatments. This may bias the results.

Authors should indicate the loss of participants for each intervention, separately for each sequence and period, possibly within the flow diagram with reasons if possible.

There are statistical methods to deal with incomplete data (see Item 12a).

Item 15: Baseline data

A table showing baseline demographic and clinical characteristics by sequence and period.

Standard CONSORT item: a table showing baseline demographic and clinical characteristics for each group

Examples

Example 1: By sequence only [43]

Characteristic	Montelukast-salmeterol (n=78)	Salmeterol-montelukast (n=76)
Sex, No (%) Male Female	 43 (55.1) 35 (44.9)	 46 (60.5) 30 (39.5)
Race, No. (%) Asian Black White Other	 1 (1.3) 11 (4.1) 38 (48.7) 28 (35.9)	 0 (0.0) 7 (9.2) 41 (53.9) 28 (36.8)
Age, mean (SD), y	10.2 (2.0)	9.8 (2.0)
Preexercise FEV_1, mean (SD), L	2.30 (1.1)	2.2 (0.6)
Preexercise FEV_1, mean (SD), % predicted	96.3 (31.8)	92.8 (12.4)
Maximum percentage decrease in FEV_1 after exercise, mean (SD)	24.8 (10.3)	25.4 (9.0)
$AUC_{0–20min}$, mean (SD), %·min	320.1 (208.6)	317.7 (165.7)
Time to recovery, mean (SD), min	23.5 (10.5)	21.5 (8.3)
Maximum FEV1, mean (SD), % predicted	99.9 (32.5)	100.5 (15.6)
Average percentage change in FEV1 after first SABA use, mean (SD)	1.4 (11.0)	4.8 (10.9)
Need for rescue medication after challenge, No. (%) No Yes	 77 (98.7) 1 (1.3)	 75 (98.7) 1 (1.3)
Asthma exacerbations limit normal physical activity, No. (%) Not at all		

Slightly	2 (2.6)	4 (5.3)
Moderately	21 (26.9)	20 (26.3)
Severely	46 (59.0)	44 (57.9)
	9 (11.5)	8 (10.5)

Abbreviations: AUC0–20min, area under the curve for the first 20 minutes after exercise; FEV1, forced expiratory volume in 1 second; SABA, short-acting β-agonist.
a Based on the number of patients who returned to within 5% of the baseline FEV1 value.

Example 2: By sequence and by total (Adapted from [55])

Characteristic	Treatment sequence		
	100 IU kg^{-1} once weekly to 50 IU kg^{-1} twice weekly (n = 22)	50 IU kg^{-1} twice weekly to 100 IU kg^{-1} once weekly (n = 25)	Total (N = 50)*
Mean age, years (SD)	31.7 (13.4)	25.1 (14.4)	27.7 (13.9)
Sex, n (%)			
Male	22 (100.0)	25 (100.0)	50 (100.0)
Ethnicity, n (%)			
White	21 (95.5)	25 (100.0)	49 (98.0)
Black	1 (4.5)	0	1 (2.0)
Hispanic or Latino	5 (22.7)	2 (8.0)	7 (14.0)
Non-Hispanic or non-Latino	17 (77.3)	23 (92.0)	43 (86.0)
Mean weight, kg (SD)	72.3 (14.2)	64.6 (26.0)	69.2 (21.3)
Target joints†, n (%)	20 (90.9)	19 (76.0)	42 (84.0)
Haemophilic arthropathy†, n (%)	20 (90.9)	17 (68.0)	40 (80.0)
Decreased movement due to haemophilic arthropathy†, n (%)	18 (81.8)	14 (56.0)	34 (68.0)

*Includes three subjects who received study drug in first on-demand period, but were not randomized.
†At study entry.
SD, standard deviation.

Explanation

Random assignment by individual ensures that any differences in group characteristics at baseline are the result of chance rather than some systematic bias [2]. For randomised crossover trials, it is desirable to know that baseline characteristics that can be affected by the intervention return to their initial state at the beginning of the second period. The by-sequence information is needed to assess whether randomisation has achieved balance

between the sequences for important variables at the start of the trial. The by-period information is helpful for readers to understand whether the treatment effect in the next period is confounded by the changing participant characteristics between periods. Characteristics that remain the same at the start of the two periods such as sex, age, for example, can be presented once; however, unstable prognostic factors and baseline value of the main outcome must be checked at beginning of each period. If the characteristic can change over time, then a baseline table by sequence only precludes inference of differences between period (i.e. treatment).

Item 16: Numbers analysed

Number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups.

Standard CONSORT item: For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups.

Explanation

The number of participants who contribute to the analysis of a trial is essential to interpreting the results. The analysis of crossover trials has to account for the paired nature of the design, the numbers analysed for each outcome should be equal to the numbers of within-participant differences or contrasts that were possible. However, not all participants may contribute to the analysis of each outcome. In a crossover trial when participants do not contribute to the analysis from one period the corresponding period may be lost. Assuming no carryover or period effect, if imputation is undertaken the data could be salvaged and when no imputation is undertaken the data is lost and becomes a power issue. As the sample size calculation and hence the power of the study is calculated on the assumption that all participants will provide information, the number of participants contributing to a particular analysis should be reported so that any potential drop in statistical power can be assessed. When there is carryover/period effect, missing data will result in a biased estimate. In addition, and as explained in detail in the CONSORT 2010 guideline [2], it should be specified whether a per-protocol or an Intention-to-treat analysis was followed.

Item 17a: Outcomes and estimation

For each primary and secondary outcome, results, including estimated effect size and its precision (such as 95% confidence interval) should be based on within participant comparisons. In addition, results for each intervention in each period are recommended.

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

Examples

Example 1: Coefficient of fat absorption (CFA) results, by treatment and severity of exocrine pancreatic insufficiency (EPI) [54]

Variable	Pancrelipase (n=16)	Placebo (n=16)	Treatment Difference (Pancrelipase-Placebo) (n=16)	P
CFA, %				
LS mean (SE)	82.8 (2.7)	47.4 (2.7)	35.4 (3.8)	<0.001
95% CI	77.0-88.6	41.6-53.2	27.2-43.6	-
CFA by severity of EPI, %				
Placebo CFA <=50%	n=10	n=10	n=10	
LS mean (SE)	81.8 (1.7)	37.3 (1.7)	44.5 (2.4)	<0.001
95% CI	77.9-85.7	33.4-41.2	39.0-50.0	-
Placebo CFA >50%	n=6	n=6	n=6	
LS mean (SE)	84.5 (2.9)	64.3 (2.9)	20.2 (4.1)	0.008
95% CI	76.5-92.5	55.3-72.3	8.9-31.6	-

Example 2: treatment comparisons and changes between baseline and treatment endpoint for secondary outcomes (Adapted from [38])

Secondary outcome	Changes between baseline to end point Mean (SE)			Treatment comparisons		
Treatment	Sildenafil PRN	Tadalafil OaD	Tadalafil PRN	LS mean difference (SE) [95% CI. P value]		
				Tadalafil OaD-Sildenafil PRN	Tadalafil OaD-Tadalafil PRN	Tadalafil PRN-Sildenafil PRN
SEAR Scale	25.40 (1.36) N = 347	25.56 (1.36) N = 348	26.92 (1.35) N = 355	0.23 (1.11) [-1.95, 2.42; *P* = 0.834]	-1.47 (1.11) [-3.65, 0.70; *P* = 0.185]	1.71 (1.10) [-0.46, 3.87; *P* = 0.123]
Sexual relationship Confidence	19.50 (1.31) N = 347	19.40 (1.31) N = 349	20.42 (1.30) N = 355	-0.07 (1.07) [-2.17, 2.04; *P* = 0.951]	-1.12 (1.06) [-3.22, 0.97; *P* = 0.291]	1.06 (1.06) [-1.03, 3.15; *P* = 0.320]
Total	22.87 (1.29) N = 347	22.94 (1.29) N = 348	24.13 (1.29) N = 355	0.11 (1.050 [-1.95, 2.17; *P* = 0.915]	-1.30 (1.040 [-3.35, 0.74; *P* = 0.212]	1.42 (1.04) [-0.63, 3.46; *P* = 0.174]
IIEF-EF Domain Score	9.70 (0.36) N = 348	8.68 (0.36) N = 350	9.54 (0.36) N = 355	-0.85 (0.30) [-1.43, -0.27; *P* = 0.004]	-0.80 (0.29) [-1.37, -0.22; *P* = 0.007]	-0.05 (0.29) [-0.62, 0.53; *P* = 0.866]
EDITS Score	75.68 (1.32) N = 348	75.81 (1.31) N = 351	79.50 (1.31) N = 355	0.12 (1.28) [-2.40, 2.64; *P* = 0.926]	-3.55 (1.27) [-6.05, -1.04; *P* = 0.006]	3.66 (1.27) [1.16, 6.17; *P* = 0.004]
Morning erection frequency	0.11 (0.02) N = 347	0.26 (0.02) N = 352	0.20 (0.02) N = 355	0.15 (0.01) [0.12, 0.18; *P* < 0.001]	0.06 (0.01) [0.03, 0.09; *P* < 0.001]	0.09 (0.01) [0.06, 0.12; *P* < 0.001]

Example 3

"Eighty patients (70%) preferred pazopanib; the most common reasons included better overall quality of life (QoL) and less fatigue. Twenty-five patients (22%) preferred sunitinib; the most common reasons included less diarrhoea and better overall QoL. Physician preferences were consistent with patient preferences. More physicians preferred to continue their patients on pazopanib (61%) than on sunitinib (22%), with 17% stating no preference." [39]

Example 4: Incidence Rate Ratios (IRR) with 95% confidence intervals for effects of treatment, time and treatment-time interaction on behaviour and affect scores after taking account of age, gender and dementia severity [56]

		Behaviour		Affect			
				Positive		Negative	
		IRR	p value	IRR	p value	IRR	p value
Treatment	Lavender compared to control	0.884 (0.778-1.004)	0.057	1.072 (0.848-1.355)	0.56	0.891 (0.504-1.573)	0.690
Time	First 30 minutes post-exposure compared to pre- exposure	0.899 (0.793-1.020)	0.097	0.900 (0.706-1.147)	0.393	0.960 (0.550-1.675)	0.887
	Second 30 minutes post-exposure compared to pre-exposure	0.858 (0.755-0.974)	0.018	0.865 (0.678-1.106)	0.248	0.641 (0.348-1.179)	0.153
Treatment-time interactions	Lavender x first 30 min post-exposure	0.961 (0.798-1.157)	0.672	1.020 (0.726-1.433)	0.910	0.848 (0.371-1.938)	0.696
	Lavender x second 30 min post-exposure	1.045 (0.869-1.259)	0.636	0.954 (0.675-1.348)	0.790	0.687 (0.269-1.750)	0.431

Explanation

When reporting the results of randomised crossover trials, point estimates with confidence intervals should be reported for primary and secondary outcomes; this is the same as the standard CONSORT guideline except that these results should be based on the appropriate within-participant analysis. Results should not be presented as though from a parallel group trial or by double counting the participants. Ideally, as the correlation

673 impacts on the power of the study, the correlation coefficient for each primary outcome
674 being analysed should also be provided to help with the planning of future crossover
675 trials.

676 For binary outcomes a presentation using a matched tabulation format is desirable as it
677 allows the reader to see the concordant and discordant pairs. The matched tabulation
678 facilitates the use of such trials in future meta-analyses as it allows using appropriate
679 formulas to adjust the between treatment variance downwards by accounting for the
680 within-participant correlation, even when not available [57-59]. Presentation of the 2x2
681 table of results from a crossover design in a parallel trial format does not allow for
682 appropriate adjustments of the between treatment variance [57]. The paired presentation
683 is also helpful for future sample size calculations. However, in many circumstances the
684 data will be analysed by a model accounting for the design and displayed as shown in
685 example 4.

686 Presentation of the results for each intervention in each period is recommended because
687 these can be used for understanding any treatment by period interaction, regardless of
688 how the trial investigators handled it in their analysis (see Table 7 of Li 2015 [31]).

689 Ideally, participant preference outcomes should also be reported at the participant level.
690 For example the participants should be split into those who prefer intervention A and
691 those who prefer B and analysed using McNemar's test or, if allowing for period, the
692 Mainland-Gart test or Prescott's test.

693

694 Item 19: Harms

695 Describe all important harms or unintended effects in a way that accounts for the design
696 (for specific guidance see CONSORT for harms [60]).

697 *Standard CONSORT item:* All important harms or unintended effects in each group (for
698 specific guidance see CONSORT for harms [60])

Examples (this example is fictional)

		Number of adverse events
Vomiting	No adverse event under either NSAIDS or placebo	108
	No adverse event under NSAIDS but adverse events observed under placebo	7
	Adverse event observed under NSAIDS but not under placebo	13
	Adverse events observed under both NSAIDS and placebo	3

Explanation

In addition to describing the types of adverse events and the overall frequency under each intervention, for crossover trials, presenting concordant and discordant pairs of adverse events or providing estimates of effect and precision (when between group comparisons were made) will inform the relative safety of the interventions tested. The table above provides an example of how to tabulate adverse events.

Discussion

Item 20: Limitations

Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses. Consider potential carry-over effects

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

Examples

Example 1

"The 24-hour washout period may have been insufficient to eliminate the effects of stimulation. Potential carryover effects should be addressed by the use of alternative

718 study designs (eg, parallel groups, longer study/washout periods, stepped-wedge
719 designs)." [45]

720 Example 2

721 "Strengths of this study include blinding of study treatments and a cross-over design,
722 where patients were exposed to both treatments in similar health states. This allowed for
723 detection of differences in tolerability not confounded by differences in health states and
724 for each patient to act as their own control. In addition, the 2-week washout period and
725 random assignment minimized possible effects of the order of treatment and carryover."
726 [39]

727 Example 3

728 "Finally, it is possible that the crossover design could have obscured differences in the
729 period on and off HCQ. While allowing for a washout period may have helped rule out
730 such a possibility, the pilot study suggested no such washout period was required." [61]

731 ***Explanation***

732 A limitation with the crossover design is that the treatment from the first period may
733 affect the results from the second period, either to improve the outcome with the
734 opposite treatment or to suppress the effect. This carry-over effect could potentially
735 render a crossover trial invalid and reporting of such a limitation is unlikely to be found
736 given that it would invalidate the trial results. Possible limitations that should be reported
737 include losses to follow-up before the second intervention is applied and mixing up of the
738 interventions such that the sequence which was applied was not that to which the
739 participant was randomised. The appropriateness of a cross-over design in terms of the
740 stability of the disease over the duration of the trial could also be discussed.
741

742

743 *More complicated trial designs*

744 In the previous sections we discussed reporting of the simple 2x2 design in which each
745 participant is randomised to one of two sequences in which to receive the two competing
746 interventions. More complicated variations of the crossover design include: comparing
747 three or more interventions (please see the CONSORT extension for multi-arm trials [62])

and cluster crossover randomised trials. In a cluster crossover randomised controlled trial, each cluster receives multiple interventions in a randomised sequence [63]. A recent review found that there is a need to ensure an appropriate analysis is undertaken and reporting needs to be improved [64]. The development of an extension of CONSORT to cluster crossover trials is underway (Joanne McKenzie, personal communication).

There also may be issues of repeated measurements (i.e. measurements taken at several timepoints) or multiplicity within participants in crossover trials (i.e. both eyes are assessed within participants). Other, less frequently used versions of the crossover design include: Bioequivalence studies, Balaam's design, extra period designs, n-of-1 designs and an incomplete block design [17].

Comment

Reports of randomised controlled trials should include key information on the methods and findings to allow readers to accurately interpret the results. This information is particularly important for meta-analysts attempting to extract data from such reports. The CONSORT 2010 statement provides the latest recommendations from the CONSORT Group on essential items to be included in the report of a randomised controlled trial. In this paper we introduce and explain corresponding updates in an extension of the CONSORT checklist specific to reporting randomised crossover trials.

Use of the CONSORT statement for the reporting of two group parallel trials is associated with improved reporting quality [65]. We believe that the routine use of this proposed extension to the CONSORT statement will eventually result in improvements to crossover designs. When reporting a randomised crossover trial, authors should address all 25 items on the CONSORT checklist using this document in conjunction with the main CONSORT guidelines [3]. Authors may also find it useful to consult the CONSORT extensions for other trial designs which are available at http://www.consort-statement.org/extensions.

The CONSORT statement can help researchers designing trials in the future and can guide peer reviewers and editors in their evaluation of manuscripts. Many journals recommend adherence to the CONSORT recommendations in their instructions to authors. We encourage them to direct authors to this and to other extensions of CONSORT for specific

trial designs. The most up to date versions of all CONSORT recommendations can be found at www.consort-statement.org.

Acknowledgements

We would like to thank James Carpenter and Mike Kenward for their input into how missing data are handled in crossover trials, and helpful comments from Sally Hopewell, Nikolaos Pandis and Drummond Rennie and other members of the CONSORT Executive, as well as the peer reviewers, Stephen Senn, Francois Curtin and Karla Hemming.

References

1. Page, M.J., et al., *Empirical Evidence of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological Studies.* PLoS One, 2016. **11**(7): p. e0159267.
2. Schulz, K.F., et al., *Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials.* JAMA, 1995. **273**(5): p. 408-12.
3. Schulz, K.F., et al., *CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials.* BMJ, 2010. **340**: p. c332.
4. Moher, D., et al., *CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials.* BMJ, 2010. **340**: p. c869.
5. Shamseer, L., et al., *Update on the endorsement of CONSORT by high impact factor journals: a survey of journal "Instructions to Authors" in 2014.* Trials, 2016. **17**(1): p. 301.
6. Campbell, M.K., et al., *Consort 2010 statement: extension to cluster randomised trials.* BMJ, 2012. **345**: p. e5661.
7. Piaggio, G., et al., *Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement.* JAMA, 2012. **308**(24): p. 2594-604.
8. Vohra, S., et al., *CONSORT extension for reporting N-of-1 trials (CENT) 2015 Statement.* BMJ, 2015. **350**: p. h1738.
9. Zwarenstein, M., et al., *Improving the reporting of pragmatic trials: an extension of the CONSORT statement.* BMJ, 2008. **337**: p. a2390.
10. Pandis, N., et al., *CONSORT 2010 statement: extension checklist for reporting within person randomised trials.* BMJ, 2017. **357**: p. j2835.
11. Zeng, L., L.A. Drye, and T. Li, *Characterizing reporting of phase 3 crossover trials registered on clinicaltrialg.gov.* Submitted, 2016.
12. Hui, D., D.S. Zhukovsky, and E. Bruera, *Which treatment is better? Ascertaining patient preferences with crossover randomized controlled trials.* J Pain Symptom Manage, 2015. **49**(3): p. 625-31.
13. Cohlen, B.J., et al., *Crossover or parallel design in infertility trials? The discussion continues.* Fertil Steril, 1998. **70**(1): p. 40-5.

14. Makubate, B. and S. Senn, *Planning and analysis of cross-over trials in infertility.* Stat Med, 2010. **29**(30): p. 3203-10.
15. Daya, S., *Differences between crossover and parallel study designs--debate?* Fertil Steril, 1999. **71**(4): p. 771-3.
16. Jones, B. and M.G. Kenward, *Design and Analysis of Cross-over Trials*. 2nd edition ed. 2003, London: Chapman&Hall/CRC.
17. Senn, S., *Crossover-trials in Clinical Research*. Second ed. 2002, Chichester: Wiley.
18. Wellek, S. and M. Blettner, *On the proper use of the crossover design in clinical trials: part 18 of a series on evaluation of scientific publications.* Dtsch Arztebl Int, 2012. **109**(15): p. 276-81.
19. Gart, J.J., *An exact test for comparing matched proportions in cross-over designs.* Biometrika, 1969. **56**: p. 75-80.
20. Mainland, D., *Elementary Medical Statistics*. 1963, W.b. Saunders: Philadelphia.
21. Freeman, P.R., *The performance of the two-stage analysis of two-treatment, two-period crossover trials.* Stat Med, 1989. **8**(12): p. 1421-32.
22. Chan, A.W. and D.G. Altman, *Epidemiology and reporting of randomised trials published in PubMed journals.* Lancet, 2005. **365**(9465): p. 1159-62.
23. Mills, E.J., et al., *Design, analysis, and presentation of crossover trials.* Trials, 2009. **10**: p. 27.
24. Yu, L.M., et al., *Reporting on covariate adjustment in randomised controlled trials before and after revision of the 2001 CONSORT statement: a literature review.* Trials, 2010. **11**: p. 59.
25. Inrig, J.K., et al., *The landscape of clinical trials in nephrology: a systematic review of Clinicaltrials.gov.* Am J Kidney Dis, 2014. **63**(5): p. 771-80.
26. Odutayo, A., et al., *Association between trial registration and positive study findings: cross sectional study (Epidemiological Study of Randomized Trials-ESORT).* BMJ, 2017. **356**: p. j917.
27. Gewandter, J.S., et al., *Reporting of cross-over clinical trials of analgesic treatments for chronic pain: Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities, and Networks systematic review and recommendations.* Pain, 2016. **157**(11): p. 2544-2551.
28. Straube, S., B. Werny, and T. Friede, *A systematic review identifies shortcomings in the reporting of crossover trials in chronic painful conditions.* J Clin Epidemiol, 2015. **68**(12): p. 1496-503.
29. Zeng, L., et al., *Registration of phase 3 crossover trials on ClinicalTrials.gov.* submitted.
30. Lathyris, D.N., T.A. Trikalinos, and J.P. Ioannidis, *Evidence from crossover trials: empirical evaluation and comparison against parallel arm trials.* Int J Epidemiol, 2007. **36**(2): p. 422-30.
31. Li, T., et al., *Design, Analysis, and Reporting of Crossover Trials for Inclusion in a Meta-Analysis.* PLoS One, 2015. **10**(8): p. e0133023.
32. Nolan, S.J., I. Hambleton, and K. Dwan, *The Use and Reporting of the Cross-Over Study Design in Clinical Trials and Systematic Reviews: A Systematic Assessment.* PLoS One, 2016. **11**(7): p. e0159014.
33. Guo, X., et al., *Effect of Ginkgo biloba on visual field and contrast sensitivity in Chinese patients with normal tension glaucoma: a randomized, crossover clinical trial.* Invest Ophthalmol Vis Sci, 2014. **55**(1): p. 110-6.

34. Johansson, D.P., et al., *Effects of unfermented and fermented whole grain rye crisp breads served as part of a standardized breakfast, on appetite and postprandial glucose and insulin responses: a randomized cross-over trial.* PLoS One, 2015. **10**(3): p. e0122241.
35. Arruda-Olson, A.M., et al., *Cardiovascular effects of sildenafil during exercise in men with known or probable coronary artery disease: a randomized crossover trial.* JAMA, 2002. **287**(6): p. 719-25.
36. Hopewell, S., et al., *CONSORT for reporting randomised trials in journal and conference abstracts.* Lancet, 2008. **371**(9609): p. 281-3.
37. Equi, A., et al., *Long term azithromycin in children with cystic fibrosis: a randomised, placebo-controlled crossover trial.* Lancet, 2002. **360**(9338): p. 978-84.
38. Rubio-Aurioles, E., et al., *A randomized open-label trial with a crossover comparison of sexual self-confidence and other treatment outcomes following tadalafil once a day vs. tadalafil or sildenafil on-demand in men with erectile dysfunction.* J Sex Med, 2012. **9**(5): p. 1418-29.
39. Escudier, B., et al., *Randomized, controlled, double-blind, cross-over trial assessing treatment preference for pazopanib versus sunitinib in patients with metastatic renal cell carcinoma: PISCES Study.* J Clin Oncol, 2014. **32**(14): p. 1412-8.
40. Konstas, A.G., et al., *24-Hour control with a latanoprost-timolol fixed combination vs timolol alone.* Arch Ophthalmol, 2006. **124**(11): p. 1553-7.
41. Gelderblom, H., et al., *Bupropion for the treatment of apathy in Huntington's disease: A multicenter, randomised, double-blind, placebo-controlled, prospective crossover trial.* PLoS One, 2017. **12**(3): p. e0173872.
42. Julious, S.A., M.J. Campbell, and D.G. Altman, *Estimating sample sizes for continuous, binary, and ordinal outcomes in paired comparisons: practical hints.* J Biopharm Stat, 1999. **9**(2): p. 241-51.
43. Fogel, R.B., et al., *Effect of montelukast or salmeterol added to inhaled fluticasone on exercise-induced bronchoconstriction in children.* Ann Allergy Asthma Immunol, 2010. **104**(6): p. 511-7.
44. Freemantle, N., et al., *Final results of the DAPS (Denosumab Adherence Preference Satisfaction) study: a 24-month, randomized, crossover comparison with alendronate in postmenopausal women.* Osteoporos Int, 2012. **23**(1): p. 317-26.
45. Abell, T.L., et al., *A double-masked, randomized, placebo-controlled trial of temporary endoscopic mucosal gastric electrical stimulation for gastroparesis.* Gastrointest Endosc, 2011. **74**(3): p. 496-503 e3.
46. Hill, J., et al., *A double-blind crossover study to compare lysine acetyl salicylate (Aspergesic) with ibuprofen in the treatment of rheumatoid arthritis.* J Clin Pharm Ther, 1990. **15**(3): p. 205-11.
47. Bonten, T.N., et al., *Time-dependent effects of aspirin on blood pressure and morning platelet reactivity: a randomized cross-over trial.* Hypertension, 2015. **65**(4): p. 743-50.
48. Fleiss, J.L., S. Wallenstein, and R. Rosenfeld, *Adjusting for baseline measurements in the two-period crossover study: a cautionary note.* Control Clin Trials, 1985. **6**(3): p. 192-7.
49. Willan, A.R. and J.L. Pater, *Using baseline measurements in the two-period crossover clinical trial.* Control Clin Trials, 1986. **7**(4): p. 282-9.

50. Carpenter, J.R. and M.G. Kenward, *Sensitivity Analysis with Multiple Imputation* in *Handbook of Missing Data Methodology*, G. Molenberghs, et al., Editors. 2015, CRC Press New York. p. 446.
51. Chen, C.Y., et al., *Effect of almond consumption on vascular function in patients with coronary artery disease: a randomized, controlled, cross-over trial.* Nutr J, 2015. **14**: p. 61.
52. Marchetti, E., et al., *Efficacy of essential oil mouthwash with and without alcohol: a 3-day plaque accumulation model.* Trials, 2011. **12**: p. 262.
53. Markman, J.D., et al., *Double-blind, randomized, controlled, crossover trial of pregabalin for neurogenic claudication.* Neurology, 2015. **84**(3): p. 265-72.
54. Graff, G.R., et al., *Efficacy and tolerability of a new formulation of pancrelipase delayed-release capsules in children aged 7 to 11 years with exocrine pancreatic insufficiency and cystic fibrosis: a multicenter, randomized, double-blind, placebo-controlled, two-period crossover, superiority study.* Clin Ther, 2010. **32**(1): p. 89-103.
55. Valentino, L.A., et al., *Multicentre, randomized, open-label study of on-demand treatment with two prophylaxis regimens of recombinant coagulation factor IX in haemophilia B subjects.* Haemophilia, 2014. **20**(3): p. 398-406.
56. O'Connor, D.W., et al., *A randomized, controlled cross-over trial of dermally-applied lavender (Lavandula angustifolia) oil as a treatment of agitated behaviour in dementia.* BMC Complement Altern Med, 2013. **13**: p. 315.
57. Elbourne, D.R., et al., *Meta-analyses involving cross-over trials: methodological issues.* Int J Epidemiol, 2002. **31**(1): p. 140-9.
58. Stedman, M.R., et al., *Meta-analyses involving cross-over trials: methodological issues.* Int J Epidemiol, 2011. **40**(6): p. 1732-4.
59. Curtin, F., D. Elbourne, and D.G. Altman, *Meta-analysis combining parallel and cross-over clinical trials. II: Binary outcomes.* Stat Med, 2002. **21**(15): p. 2145-59.
60. Ioannidis, J.P., et al., *Better reporting of harms in randomized trials: an extension of the CONSORT statement.* Ann Intern Med, 2004. **141**(10): p. 781-8.
61. Solomon, D.H., et al., *Effect of hydroxychloroquine on insulin sensitivity and lipid parameters in rheumatoid arthritis patients without diabetes mellitus: a randomized, blinded crossover trial.* Arthritis Care Res (Hoboken), 2014. **66**(8): p. 1246-51.
62. Juszczak, E., et al., *CONSORT statement: extension to multi-arm parallel RCTs.* submitted.
63. Rietbergen, C. and M. Moerbeck, *The Design of Cluster Randomized Crossover Trials.* Journal of Educational and Behavioral Statistics, 2011. **36**(4): p. 472-490.
64. Arnup, S.J., et al., *Appropriate statistical methods were infrequently used in cluster-randomized crossover trials.* J Clin Epidemiol, 2016. **74**: p. 40-50.
65. Turner, L., et al., *Does use of the CONSORT Statement impact the completeness of reporting of randomised controlled trials published in medical journals? A Cochrane review.* Syst Rev, 2012. **1**: p. 60.

Note that small parts of the text in this manuscript are necessarily similar to other CONSORT articles.