Reporting of Noninferiority and Equivalence Randomized Trials
Extension of the CONSORT 2010 Statement

Gilda Piaggio, PhD
Diana R. Elbourne, PhD
Stuart J. Pocock, PhD
Stephen J. W. Evans, MSc
Douglas G. Altman, DSc
for the CONSORT Group

The CONSORT (Consolidated Standards of Reporting Trials) Statement, which includes a checklist and a flow diagram, is a guideline developed to help authors improve the reporting of the findings from randomized controlled trials. It was updated most recently in 2010. Its primary focus is on individually randomized trials with 2 parallel groups that assess the possible superiority of one treatment compared with another. The CONSORT Statement has been extended to other trial designs such as cluster randomization, and recommendations for noninferiority and equivalence trials were made in 2006. In this article, we present an updated extension of the CONSORT checklist for reporting noninferiority and equivalence trials, based on the 2010 version of the CONSORT Statement and the 2008 CONSORT Statement for the reporting of abstracts, and provide illustrative examples and explanations for those items that differ from the main 2010 CONSORT checklist. The intent is to improve reporting of noninferiority and equivalence trials, enabling readers to assess the reliability of their results and conclusions.

JAMA. 2012;308(24):2594-2604 www.jama.com

See also page 2605.

©2012 American Medical Association. All rights reserved.
Because proof of exact equivalence is impossible, a pretest margin of non-inferiority (\(\Delta\)) for the treatment effect in a primary patient outcome is defined. Equivalence trials are very similar, except that equivalence is defined as the treatment effect being between \(-\Delta\) and \(+\Delta\). For therapeutic or prophylactic trials the noninferiority approach is much more common than a true (2-sided) equivalence approach. However, equivalence trials are more common in pharmacokinetics, in which a difference in either direction from the reference treatment is of importance.

Noninferiority of the new treatment with respect to the reference treatment is of interest on the premise that the new treatment has some other advantage, such as greater availability, reduced cost, less invasiveness, \(^{11,12}\) fewer adverse effects (harms), \(^{13}\) or greater ease of administration. \(^{14}\) In trials that investigate noninferiority, therefore, the question of interest is not symmetric. \(^{15}\) The new treatment will be recommended if it is similar to the reference treatment for a prespecified primary outcome but not if it is worse by more than \(\Delta\). Superiority of the new treatment for the primary outcome would be an additional benefit. Some noninferiority trials have been criticized for merely studying a new marketable product (“me-too” drugs) without offering any advantages over existing products. \(^{16}\) The use of noninferiority or equivalence trials has been criticized on the grounds that they ask “no relevant clinical questions” and are therefore unethical. \(^{17}\) But some observers argue that this view is misplaced. \(^{18,19}\)

This article focuses mainly on noninferiority trials but applies also to the less common 2-sided equivalence trials (eAppendix, available at http://www.jama.com).

How Common Are Noninferiority Trials?
Assessing the frequency of noninferiority trials is complicated, because not all noninferiority or equivalence trials use these words, and the term “equivalence” is often inappropriately used when reporting “negative” (null) results of superiority trials; such trials often lack statistical power to rule out important differences. \(^{20,21}\)

A recent review of 583 noninferiority trials of drug therapies published between 1989 and 2009 showed an increasing trend, with only 1 trial published before 1999 and more than 100 trials published per year from 2007. \(^{22}\) A third of these were in the fields of infectious diseases or cardiology. An earlier review found the same 2 specialties had the greatest number of noninferiority and equivalence trials. \(^{23}\) Surveys in ophthalmology \(^{24}\) and oncology \(^{25}\) also found increases in the number of such trials.

Quality of Reporting of Noninferiority Trials
Early reviews of the quality of trials claiming equivalence found that important deficiencies were common. Equivalence was inappropriately claimed in 67% of 88 studies published from 1992 to 1996 on the basis of nonsignificant tests for superiority. \(^{26}\) Fifty-one percent stated equivalence as an aim, but only 23% reported that they were designed with a preset margin of equivalence. Other disease- or field-specific reviews had similar findings. \(^{27-29}\)

More recent reviews have found that the quality of reports of noninferiority and equivalence trials remains poor. In one review (covering the years 1990 to 2000) only about one-fifth of 332 noninferiority and equivalence trials provided a suitable rationale for the noninferiority margin. \(^{30}\) In another review covering noninferiority trials indexed in PubMed as of February 5, 2009, almost all of 232 published reports of equivalence and noninferiority drug trials specified the noninferiority margin, but only 24% explained how it was determined. \(^{30-31}\) Other reviews had broadly similar findings. \(^{32-34}\) An increasing quality of reporting of noninferiority trials in oncology was observed from 2001 to 2010. \(^{35}\)

Box. Major Changes From the 2006 Extension of the CONSORT Statement for Reporting Noninferiority and Equivalence Trials
- Based on the standard CONSORT 2010 checklist, which incorporates changes to the CONSORT 2001 checklist described in detail in the CONSORT 2010 Statement \(^{9}\)
- Uses a 2-column display to show more clearly the additional information to report for noninferiority trials
- New checklist for abstracts to apply to noninferiority trials
- Expanded checklist items for objectives, outcomes, and interpretation
- Most examples of good reporting practice updated, including 10 new examples of good reporting from publications after 2006. Kept 3 examples from the 2006 extension that illustrate specific points
- Recent methodological developments are summarized in the eAppendix
- Empirical evidence of reporting of noninferiority trials updated

Updating the CONSORT Statement Extension for Noninferiority Randomized Trials
The updated CONSORT 2010 Statement comprises a 25-item checklist and a participant flow diagram. \(^{4}\) In the 2010 update, some new items and subitems were introduced, wording was simplified and clarified, and the specificity of some items was made more explicit by breaking them into subitems. Methodological advances reported in the literature since the 2001 Statement were reviewed and taken into consideration. This noninferiority extension was undertaken to reflect the updated CONSORT Statement and to integrate any significant advances in noninferiority trials methodology since 2006.
The Updating Process

An electronic search of publications citing the original CONSORT extension for noninferiority and equivalence trials was conducted using Web of Science (October 14, 2010). The search yielded 260 publications. An initial assessment of the titles and abstracts was made for relevance, yielding 142 articles. After excluding repeated publications, 137 articles remained, of which 85 were trial reports, 47 were methodological papers, and 5 were reviews of published reports of trials potentially relevant to the update of the CONSORT extension. The methodological studies and reviews were assessed for material that might influence the update. In addition, we reviewed publications from 2006 and later, including guidelines issued by both the Food and Drug Administration and the European Medicines Agency.

Figure 1. Possible Scenarios of Observed Treatment Differences for Adverse Outcomes (Harms) in Noninferiority Trials

Error bars indicate 2-sided 95% CIs. The blue dashed line at $x = \Delta$ indicates the noninferiority margin; the blue tinted region to the left of $x = \Delta$ indicates the zone of inferiority. A, If the CI lies wholly to the left of zero, the new treatment is superior. B and C, If the CI lies to the left of $\Delta$ and includes zero, the new treatment is noninferior but not shown to be superior. D, If the CI lies wholly to the left of $\Delta$ and wholly to the right of zero, the new treatment is noninferior in the sense already defined but also inferior in the sense that a null treatment difference is excluded. This puzzling circumstance is rare, because it requires a very large sample size. It also can result from a noninferiority margin that is too wide. E and F, If the CI includes $\Delta$ and zero, the difference is nonsignificant but the result regarding noninferiority is inconclusive. G, If the CI includes $\Delta$ and is wholly to the right of zero, the difference is statistically significant but the result is inconclusive regarding possible inferiority of magnitude $\Delta$ or worse. H, If the CI is wholly above $\Delta$, the new treatment is inferior.

Methodology

Methodological considerations in noninferiority trials are discussed in the eAppendix. Key issues include the need to state the trial hypotheses in relation to the noninferiority margin; the choice of margin; analysis using a CI approach; and the presentation and interpretation of the results using the CI in relation to the noninferiority margin (Figure 1).

Revised Extension of CONSORT Statement

To accommodate noninferiority trials, an extension of the CONSORT Statement should encompass the following main issues: (1) the rationale for adopting a noninferiority design; (2) how study hypotheses were incorporated into the design; (3) choice of participants, interventions (especially the reference treatment), and outcomes; (4) statistical methods, including sample size calculation; and (5) how the design affects interpretation and conclusions. Consequences for the CONSORT checklist, including specific changes, are described below. The flow diagram was not considered to require any specific modification.

Checklist

The revised checklist for the reporting of noninferiority trials, updated in line with the CONSORT 2010 Statement, is presented in Table 1. This checklist relates to noninferiority trials, but the same issues apply to equivalence trials.

We have reformatted the checklist in line with the style currently promoted by the CONSORT Group, as used for the extensions for nonpharmacological interventions, pragmatic trials, and cluster randomized trials. The text in 2 columns, the first comprising the CONSORT 2010 checklist and the second the revised extension for noninferiority trials. Several items are extended to cover reporting recommendations specific to the noninferiority design.

For each extended item, we include 1 or more examples of good reporting and...
Table 1. Information to Include When Reporting a Noninferiority or Equivalence Randomized Trial: Extension of CONSORT 2010 Checklist

<table>
<thead>
<tr>
<th>Section/Topic</th>
<th>Item No.</th>
<th>Standard CONSORT 2010 Checklist Item</th>
<th>Extension for Noninferiority Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title and Abstract</td>
<td>1a</td>
<td>Identification as a randomized trial in the title</td>
<td>Identification as a noninferiority randomized trial in the title</td>
</tr>
<tr>
<td>Abstract</td>
<td>1b</td>
<td>Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)(7)(8)</td>
<td>See Table 2</td>
</tr>
<tr>
<td>Introduction</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Background and objectives</td>
<td>2a</td>
<td>Scientific background and explanation of rationale</td>
<td>Rationale for using a noninferiority design</td>
</tr>
<tr>
<td></td>
<td>2b</td>
<td>Specific objectives or hypotheses</td>
<td>Hypotheses concerning noninferiority, specifying the noninferiority margin with the rationale for its choice</td>
</tr>
<tr>
<td>Methods</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial design</td>
<td>3a</td>
<td>Description of trial design (such as parallel, factorial), including allocation ratio</td>
<td>Whether the reference treatment in the noninferiority trial is identical (or very similar) to that in any trial(s) that established efficacy</td>
</tr>
<tr>
<td></td>
<td>3b</td>
<td>Important changes to methods after trial commencement (such as eligibility criteria), with reasons</td>
<td>Whether participants in the noninferiority trial are similar to those in any trial(s) that established efficacy of the reference treatment</td>
</tr>
<tr>
<td>Participants</td>
<td>4a</td>
<td>Eligibility criteria for participants</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4b</td>
<td>Settings and locations where the data were collected</td>
<td></td>
</tr>
<tr>
<td>Interventions</td>
<td>5</td>
<td>The interventions for each group with sufficient details to allow replication, including how and when they were actually administered</td>
<td>Whether the reference treatment in the noninferiority trial is identical (or very similar) to that in any trial(s) that established efficacy</td>
</tr>
<tr>
<td>Outcomes</td>
<td>6a</td>
<td>Completely defined prespecified primary and secondary outcome measures, including how and when they were assessed</td>
<td>Specify the noninferiority outcome(s) and whether hypotheses for main and secondary outcome(s) are noninferiority or superiority. Whether the outcomes in the noninferiority trial are identical (or very similar) to those in any trial(s) that established efficacy of the reference treatment</td>
</tr>
<tr>
<td></td>
<td>6b</td>
<td>Any changes to trial outcomes after the trial commenced, with reasons</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td>7a</td>
<td>How sample size was determined</td>
<td>Whether the sample size was calculated using a noninferiority criterion and, if so, what the noninferiority margin was</td>
</tr>
<tr>
<td></td>
<td>7b</td>
<td>When applicable, explanation of any interim analyses and stopping guidelines</td>
<td>To which outcome(s) they apply and whether related to a noninferiority hypothesis</td>
</tr>
<tr>
<td>Sequence generation</td>
<td>8a</td>
<td>Method used to generate the random allocation sequence</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8b</td>
<td>Type of randomization; details of any restriction (such as blocking and block size)</td>
<td></td>
</tr>
<tr>
<td>Allocation concealment mechanism</td>
<td>9</td>
<td>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</td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td>10</td>
<td>Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions</td>
<td></td>
</tr>
<tr>
<td>Blinding</td>
<td>11a</td>
<td>If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11b</td>
<td>If relevant, description of the similarity of interventions</td>
<td></td>
</tr>
<tr>
<td>Statistical methods</td>
<td>12a</td>
<td>Statistical methods used to compare groups for primary and secondary outcomes</td>
<td>Whether a 1- or 2-sided confidence interval approach was used</td>
</tr>
<tr>
<td></td>
<td>12b</td>
<td>Methods for additional analyses, such as subgroup analyses and adjusted analyses</td>
<td></td>
</tr>
<tr>
<td>Participant flow (a diagram is strongly recommended)</td>
<td>13a</td>
<td>For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analyzed for the primary outcome</td>
<td></td>
</tr>
<tr>
<td></td>
<td>13b</td>
<td>For each group, losses and exclusions after randomization, together with reasons</td>
<td></td>
</tr>
<tr>
<td>Recruitment</td>
<td>14a</td>
<td>Dates defining the periods of recruitment and follow-up</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14b</td>
<td>Why the trial ended or was stopped</td>
<td></td>
</tr>
<tr>
<td>Baseline data</td>
<td>15</td>
<td>A table showing baseline demographic and clinical characteristics for each group</td>
<td></td>
</tr>
<tr>
<td>No. analyzed</td>
<td>16</td>
<td>For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups</td>
<td></td>
</tr>
<tr>
<td>Outcomes and estimation</td>
<td>17a</td>
<td>For each primary and secondary outcome, results for each group, the estimated effect size, and its precision (such as 95% confidence interval)</td>
<td>For the outcome(s) for which noninferiority was hypothesized, a figure showing confidence intervals and the noninferiority margin may be useful</td>
</tr>
<tr>
<td></td>
<td>17b</td>
<td>For binary outcomes, presentation of both absolute and relative effect sizes is recommended</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
provide explanatory text. In some of the examples we have added text in brackets to explain the context. In some cases, a particular item is well reported, but providing an example does not imply that all aspects are well reported. For some items it was not possible to find a perfectly reported example. Throughout the literature, authors use different pairs of comparative terms (eg, greater than/less than, better/worse) to characterize the direction of effects, depending on whether the end points are positive (eg, survival) or negative (eg, adverse events). We have not changed the original text of each example but tried to clarify the meaning of the comparative terms used, where it might be confusing.

**Title and Abstract**

**Item 1a: Title.** Standard CONSORT item: Identification as a randomized trial in the title. *Extension for noninferiority trials:* Identification as a noninferiority randomized trial in the title.

**Example.** “Dabigatran Etxetilate Versus Enoxaparin for Prevention of Venous Thromboembolism After Total Hip Replacement: A Randomised, Double-Blind, Non-Inferiority Trial.”

**Explanation.** Readers should be able to easily identify from the title or abstract that the study was a noninferiority or equivalence randomized trial. Including the design in the title or abstract also ensures ease of identification of these studies in a literature search for inclusion in systematic reviews.

**Item 1b: Abstract.** Standard CONSORT item: Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts). *Extension for noninferiority trials: See Table 2.*

**Example.** This example details only those parts relevant to noninferiority.

**Title: Identification of Study as a Noninferiority Trial.** “Duloxetine, Pregabalin, and Duloxetine plus Gabapentin for Diabetic Peripheral Neuropathic Pain Management in Patients With Inadequate Pain Response to Gabapentin: An Open-Label, Randomized, Noninferiority Comparison.”

**Methods-Objective: Specific Hypothesis Concerning Noninferiority, Including Noninferiority Margin.** “To determine whether duloxetine is noninferior to (as good as) pregabalin in the treatment of pain associated with diabetic peripheral neuropathy. . . . Noninferiority would be declared if the mean improvement [in the weekly mean of the diary-based daily pain score] for duloxetine was no worse than the mean improvement for pregabalin, within statistical variability, by a margin of −0.8 unit.”

**Methods-Outcome: Clarify for all reported outcomes whether noninferiority or superiority.** “The primary objective was a noninferiority comparison between duloxetine and pregabalin on improvement in the weekly mean of the diary-based daily pain score (0- to 10-point scale) at end point.

“. . . adverse effects, nausea, insomnia, hyperhidrosis, and decreased appetite [were secondary outcomes to be assessed for superiority].”

**Results-Outcome: For the Primary Noninferiority Outcome, Results in Relation to Noninferiority Margin.** “The 97.5% lower confidence limit was a −0.05 difference in means, establishing noninferiority.”

**Conclusions: Interpretation Taking into Account the Noninferiority Hypotheses and Any Superiority Hypotheses.** “Duloxetine was noninferior to pregabalin for the treatment of pain in patients with diabetic peripheral neuropathy who had an inadequate pain response to gabapentin.”

**Explanation.** Clear, transparent, and sufficiently detailed abstracts are important. Readers may only have access to the abstract, and many others skim it before deciding whether to read further. A well-written abstract also helps in retrieval of relevant reports from elec-

---

### Table 1. Information to Include When Reporting a Noninferiority or Equivalence Randomized Trial: Extension of CONSORT 2010 Checklist

<table>
<thead>
<tr>
<th>Section/Topic</th>
<th>Item No.</th>
<th>Standard CONSORT 2010 Checklist Item</th>
<th>Extension for Noninferiority Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ancillary analyses</td>
<td>18</td>
<td>Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing prespecified from exploratory</td>
<td></td>
</tr>
<tr>
<td>Harms</td>
<td>19</td>
<td>All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)</td>
<td></td>
</tr>
<tr>
<td>Limitations</td>
<td>20</td>
<td>Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses</td>
<td></td>
</tr>
<tr>
<td>Generalizability</td>
<td>21</td>
<td>Generalizability (external validity, applicability) of the trial findings</td>
<td></td>
</tr>
<tr>
<td>Interpretation</td>
<td>22</td>
<td>Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence Interpret results in relation to the noninferiority hypothesis. If a superiority conclusion is drawn for outcome(s) for which noninferiority was hypothesized, provide justification for switching</td>
<td></td>
</tr>
<tr>
<td>Registration</td>
<td>23</td>
<td>Registration number and name of trial registry</td>
<td></td>
</tr>
<tr>
<td>Protocol</td>
<td>24</td>
<td>Where the full trial protocol can be accessed, if available</td>
<td></td>
</tr>
<tr>
<td>Funding</td>
<td>25</td>
<td>Sources of funding and other support (such as supply of drugs), role of funders</td>
<td></td>
</tr>
</tbody>
</table>

*This checklist relates to noninferiority trials, but the same issues apply to equivalence trials. The Consolidated Standards of Reporting Trials (CONSORT) Group “strongly recommends reading this checklist in conjunction with the CONSORT 2010 Explanation and Elaboration for important clarifications on all the items.” This checklist may be republished without restriction.*

©2012 American Medical Association. All rights reserved.
tronic databases. In 2008, a CONSORT extension for reporting abstracts was published,7,8 and those recommendations were incorporated into CONSORT 2010. For noninferiority studies, the study design24 and the noninferiority margin12 are poorly reported in abstracts. In addition to the items recommended for all trials, abstracts for noninferiority RCTs should specify the noninferiority hypothesis, identify the primary outcome and noninferiority margin, and make clear whether hypotheses for other reported outcomes are noninferiority or superiority. The results should relate the primary noninferiority outcome to the noninferiority margin. The overall interpretation should take account of noninferiority and also any superiority hypotheses (Table 2).

**Introduction**

**Item 2a: Background.** Standard CONSORT item: Scientific background and explanation of rationale. Extension for noninferiority trials: Rationale for using a noninferiority design.

Example. “German guidelines consider adjuvant fluorouracil the standard of care [for locally advanced rectal cancer]. Optimisation of local tumour control has meant that distant metastases now represent the most common type of treatment failure in rectal cancer. . . . Capecitabine is an oral fluoropyrimidine derivative that was as effective as fluorouracil plus folinic acid for adjuvant treatment of stage III colon cancer. It was also non-inferior to infusional fluorouracil in combination with oxaliplatin for first-line treatment of metastatic colorectal cancer. . . .”

**Explanation.** The rationale for using a noninferiority design should include evidence for the efficacy of the reference treatment in a similar context. If previous trials (preferably part of a systematic review) demonstrated the superiority of the reference treatment relative to placebo (or an equivalent, such as “usual care” for nonpharmacological interventions) they should be cited, preferably with effect sizes and CIs. If no such trials exist, other evidence for efficacy of the reference treatment should be given. Evidence for other potential advantages of the new treatment over the reference treatment should be summarized, to justify use of the new treatment if it should be shown to be noninferior. One aim of the current trial might be to provide or support such evidence.

(See also checklist items 4a, 5, and 6.)

**Item 2b: Objectives.** Standard CONSORT item: Specific objectives or hypotheses. Extension for noninferiority trials: Hypotheses concerning noninferiority, specifying the noninferiority margin with the rationale for its choice.

Example. “A sequential analysis for the antiplatelet comparison was developed and planned to first test the noninferiority of aspirin plus extended-release dipyridamole as compared with clopidogrel. If this condition was sat-

| Table 2. Information to Include in the Abstract of a Report of a Noninferiority or Equivalence Randomized Trial: Extension of CONSORT for Abstracts Checklist*ª,² |
|-----------------|-----------------|-----------------|
| Item | Standard Checklist Item | Extension for Noninferiority Trials |
| Title | Identification of study as randomized | Identification of study as a noninferiority trial |
| Trial design | Description of the trial design (eg, parallel, cluster, noninferiority) | |
| Methods | | |
| Participants | Eligibility criteria for participants and the settings in which the data were collected | |
| Interventions | Interventions intended for each group | Specific hypothesis concerning noninferiority, including noninferiority margin |
| Objective | Specific objective or hypothesis | |
| Outcome | Clearly defined primary outcome for this report | |
| Randomization | How participants were allocated to interventions | |
| Blinding (masking) | Whether participants, caregivers, and those assessing the outcomes were blinded to group assignment | |
| Numbers randomized | Number of participants randomized to each group | |
| Numbers analyzed | Number of participants analyzed in each group | |
| Results | For the primary outcome, a result for each group and the estimated effect size and its precision | For the primary noninferiority outcome, results in relation to noninferiority margin |
| Recruitment | Trial status | |
| Harms | Important adverse events or side effects | |
| Conclusions | General interpretation of the results | Interpretation taking into account the noninferiority hypotheses and any superiority hypotheses |
| Trial registration | Registration number and name of trial register | |
| Funding | Source of funding | |

Abbreviation: CONSORT, Consolidated Standards of Reporting Trials.

*This checklist relates to noninferiority trials, but the same issues apply to equivalence trials.

ªThis checklist may be republished without restriction.
isfied, then the superiority of aspirin plus extended-release dipyridamole over clopidogrel could be assessed in a second test of the conventional null hypothesis of no difference between the two treatments.

Confirmation of noninferiority in this trial involved the prespecification of a hazard ratio for aspirin plus extended-release dipyridamole, as compared with clopidogrel, that is below a predefined margin. The margin was defined in the following way. . . . "45 (See item 7a.)

Explanation. The authors should specify for which outcomes noninferiority hypotheses apply and for which superiority hypotheses apply. Usually the noninferiority hypothesis refers to the primary end point, whereas the new treatment is expected to offer other advantages, eg, fewer adverse effects or lower cost. If the trial is multigroup or the treatments have a factorial structure, the comparisons to which the noninferiority hypothesis applies should be specified. If sequential testing of noninferiority and superiority hypotheses was planned, that should also be reported.

The rationale for the choice of the noninferiority margin and whether the margin is based on a relative or absolute scale should be specified because relative measures tend to make it less easy to conclude noninferiority, particularly when observed rates turn out to be smaller than the expected rates.46,47

The method used to set the margin of noninferiority should be reported. Conventionally, the margin is taken as the size of the effect considered clinically irrelevant. That approach might show an ineffective new treatment as noninferior if the margin is too large in relation to the effect of the reference treatment compared with placebo. To prove that the new treatment is effective, the effect retention or putative placebo effect (eAppendix),36 and it should be used if possible if the noninferiority trial is aimed for drug approval.35,47

METHODS

Item 4a: Participants. Standard CONSORT item: Eligibility criteria for participants. Extension for noninferiority trials: Whether participants in the noninferiority trial are similar to those in any trial(s) that established efficacy of the reference treatment.

Example. “[We] enrolled 6628 men and women in 312 health centres in Sweden . . . who had hypertension (blood pressure ≥180 mm Hg systolic, ≥105 mm Hg diastolic, or both), aged 70-84 years. The only difference in inclusion criteria between this trial and the STOP-Hypertension trial was that patients with isolated systolic hypertension could be included in STOP Hypertension-2, based on previous positive findings in patients with isolated systolic hypertension treated with diuretics and calcium antagonists.”48

Explanation. Because an inference of noninferiority relies on evidence that the reference treatment is effective (see “Assay Sensitivity” in eAppendix), relevant differences in participants’ characteristics compared with previous trials should be reported and explained. Such description should concentrate on differences that might affect response to treatments. For continuous variables it is important to provide not just the mean values but also an indication of variability (eg, standard deviation).

Item 5: Interventions. Standard CONSORT item: The interventions for each group with sufficient details to allow replication, including how and when they were actually administered. Extension for noninferiority trials: Whether the reference treatment in the noninferiority trial is identical (or very similar) to that in any trial(s) that established efficacy.

Example. “[S]even large, randomised, placebo-controlled trials involving a total of 16,770 patients who underwent percutaneous interventions have established that the overall reduction in the risk of death or nonfatal myocardial infarction 30 days after adjunctive inhibition of platelet glycoprotein IIb/IIIa receptors is 38 percent [relative reduction]. . . . The primary end point [in the present trial] was a composite of death, nonfatal myocardial infarction, or urgent target-vessel revascularization within 30 days after the index procedure.”31

Explanation. Any differences in outcome measures in the new trial compared with trials that established efficacy of the reference treatment should be noted and explained. In particular,
authors should note any differences in the timing of evaluation. Ideally, outcomes should not be changed, but changes may be indicated by improvements in the understanding, management, and prognosis of a disease. For example, early acquired immunodeficiency syndrome (AIDS) trials had death as the primary outcome, but as deaths became uncommon, the focus shifted to AIDS clinical events, then shifted again to surrogate markers as clinical events became uncommon.

**Item 7a: Sample Size.** *Standard CONSORT item:* How sample size was determined. **Extension for noninferiority trials:** Whether the sample size was calculated using a noninferiority criterion and, if so, what the noninferiority margin was.

**Example 1 (noninferiority).** “Using data from the nonfatal stroke outcomes from the Clopidogrel versus Aspirin in Patients at Risk of Ischemic Events trial and from the meta-analysis by the Antithrombotic Trialists’ Collaboration . . . , we derived an estimated odds ratio for clopidogrel being better than placebo for the outcome of nonfatal stroke: 1.377 (95% confidence interval [CI], 1.155 to 1.645). Thus, to ensure that the aspirin plus extended-release dipyridamole preserved at least half the effect of clopidogrel, the noninferiority margin was set at 1.075, an effect size equal to half the lower limit of the confidence interval. . . . With 1715 recurrent strokes, we would have a statistical power of 82% to reject the inferiority null hypothesis, assuming a 6.5% relative risk reduction with aspirin plus extended-release dipyridamole as compared with clopidogrel.”

**Example 2 (equivalence).** “The margin of equivalence, Δ, was 5% and the range −5% to 5% was predefined as an acceptable range of completion rates [of medical abortion] between the two types of providers. The margin was based on clinically and statistically important differences as well as ethical criteria, cost, and feasibility. The sample size of 1086 women was calculated to be sufficient (with a two-sided 95% CI and 80% power) to establish equivalence. The sample size calculation allowed for 10% loss to follow-up. . . .”

**Explanation.** The margin of noninferiority Δ should be specified and preferably justified on clinical grounds. If Δ is too large, there will be too great a risk of accepting a truly inferior treatment as noninferior. This concern is especially relevant for serious outcomes such as mortality. On the other hand, defining a very small Δ might produce inconclusive results, requiring an extremely large trial if adequate power is to be achieved. If Δ is chosen to be a proportion of the difference between reference treatment and placebo in previous trials (ratio approach), that should be noted.

Calculation of power requires that the investigators stipulate the expected response in each group. It is common for these values to be set equal so that the power of the trial corresponds to the case in which there is a zero difference between the 2 groups. The power can be higher if the new treatment is assumed to be more effective than the reference treatment or lower if it is assumed to be less effective.

Two reviews of published trials found that less than three-quarters of reports of noninferiority and equivalence trials reported a sample-size calculation that incorporated Δ.

**Item 7b: Interim analyses and stopping guidelines.** *Standard CONSORT item:* When applicable, explanation of any interim analyses and stopping guidelines. **Extension for noninferiority trials:** To which outcome(s) they apply and whether related to a noninferiority hypothesis.

**Example [noninferiority trial with stopping criterion based on superiority].** “A data and safety monitoring board reviewed the data periodically for safety and efficacy. They could recommend stopping the study if a benefit in favour of oral anticoagulation therapy was shown, such that the hazard ratio for clopidogrel plus aspirin versus oral anticoagulation therapy exceeded 1.0 by more than 3 SD at either of two formal interim analyses, timed to occur when 50% or 75% of events had occurred. . . .”

**Explanation.** In superiority trials, if an interim analysis shows clear evidence of the efficacy of the new treatment, it may be considered unethical to continue the trial and deny the new effective treatment to the control group. In contrast, in noninferiority trials, if noninferiority is demonstrated for the primary outcome (using the prestated noninferiority margin) before completion of the trial, there is less ethical need to stop the trial because the control group is already receiving the standard treatment and the experimental treatment is not appearing appreciably worse. Also, if noninferiority is evident at interim analysis and the point estimate is favorable, the investigators or the data monitoring committee may then wish to continue in the hope of demonstrating superiority. In noninferiority trials it is therefore often more appropriate to base stopping rules on safety outcomes and superiority hypotheses. Stopping rules for efficacy in noninferiority trials may be asymmetric, i.e., may favor stopping early if the new treatment is appearing worse than the standard but continuing longer if the new treatment is appearing better. Formal stopping rules for futility may be particularly important for noninferiority trials (given that the comparison is with a proven standard therapy). It has been suggested that relating the observed effect to the point of “no effect” rather than the noninferiority margin may be more appropriate for considering futility and harm in noninferiority trials and that “the data would have to show convincing evidence of harm before the trial would be stopped for futility.”

**Item 12a: Statistical Methods.** *Standard CONSORT item:* Statistical methods used to compare groups for primary and secondary outcomes. **Extension for noninferiority trials:** Whether a 1- or 2-sided confidence interval approach was used.

**Example 1 (noninferiority, continuous outcome).** “The primary efficacy endpoint was the mean change in pain intensity. . . . Study endpoints were analysed primarily for the per protocol population.
and repeated, for sensitivity reasons, for the intention-to-treat (ITT) population. For most efficacy endpoints, a confidence interval (CI) approach was used on an analysis of covariance (ANCOVA) model, with a two-sided 5% level of significance. . . . For the primary efficacy endpoint, non-inferiority of lumiracoxib to indomethacin could be claimed if the lower limit of the CI [for the difference in mean change of pain intensity assessed on a 5-point Likert scale] was greater than −0.5. This test for non-inferiority was only performed for the primary efficacy variable; all other secondary variables were tests of superiority of lumiracib versus indomethacin.\(^5^n\)

**Example 2 (non-inferiority, binary outcome).** The trial was powered for separate comparisons between the control group [unfractionated heparin or enoxaparin plus a glycoprotein IIb/IIIa inhibitor] and each of the two investigational groups. We used sequential non-inferiority and superiority analyses with hierarchical end-point testing, with the type I error controlled by the Benjamini and Hochberg procedure, as previously described. Non-inferiority was declared if the upper limit of the one-sided 97.5% confidence interval (CI) for the event rate in the investigational group did not exceed a relative margin of 25% from the event rate in the control group [risk ratio = 1.25], equivalent to a one-sided test with an alpha value of 0.025. A two-sided alpha value of 0.05 was used for superiority testing.\(^5^n\)

**Example 3 (equivalence, binary outcome).** To assess the equivalence between midlevel healthcare providers and doctors, the risk difference between the two provider types together with their [two-sided] 95% CI was derived by use of a generalised estimating equation (GEE) model. . . . [The primary endpoint was complete abortion] . . . If the CI of the risk difference between the two groups falls within the predetermined margin of equivalence (−5% to 5%), the two types of providers can be considered equivalent. . . . The analyses for the primary and secondary endpoints were on an intention-to-treat basis, supplemented by per-protocol analysis of the primary endpoint.\(^5^n\)

**Explanation.** Tests of non-inferiority need to be related to the \(\Delta\) and \(\alpha\) as pre-specified in the non-inferiority hypothesis. It should be specified whether an absolute difference between treatments or a relative measure, or both, will be used. Judgment of the results in relation to the study hypothesis is based on the location of the whole CI in relation to \(\Delta\) (Figure 1). For non-inferiority trials, the upper bound of the 2-sided \((1 - 2\alpha) \times 100\%\) CI for the (deleterious) treatment effect or the upper bound of the 1-sided \((1 - \alpha) \times 100\%\) CI has to be below the margin \(\Delta\) to declare that non-inferiority has been shown, with a significance level \(\alpha\). The 2-sided CI provides additional information, in particular for the situation in which the new treatment is superior to the reference treatment. For equivalence trials, equivalence is demonstrated if the entire 2-sided \((1 - \alpha) \times 100\%\) CI lies within \(-\Delta\) and \(\Delta\).

If non-inferiority has been demonstrated, it is then acceptable to assess whether the new treatment appears superior to the reference treatment, using an appropriate test or CI, with a significance level or confidence, respectively, defined a priori in the protocol and with an ITT analysis. Conversely, occasionally a trial protocol may specify that if superiority is not demonstrated, a non-inferiority analysis will be performed.\(^6^n\) Such sequential testing should be fully explained.

**Results**

**Item 17a: Outcomes and Estimation.** Standard CONSORT item: For each primary and secondary outcome, results for each group, the estimated effect size and its precision (such as 95% CI). Extension for non-inferiority trials: For the outcome(s) for which non-inferiority was hypothesized, a figure showing CIs and the non-inferiority margin may be useful.

**Example (non-inferiority of new treatment).** The unadjusted HR comparing overall survival between the SLND [sentinel lymph node dissection]-alone group and the ALND [axillary lymph node dissection] group was 0.79 (90% CI, 0.56-1.10), which did not cross the specified boundary of 1.3. The HR for overall survival adjusting for adjuvant therapy . . . and age for the SLND-alone group compared with the ALND group was 0.87 (90% CI, 0.62-1.23) (Figure 2).\(^6^n\)

**Explanation.** A figure helps readers to interpret the result based on the CI, because it shows graphically where the CI lies with respect to the null value (if a risk difference is used) or to 1 (if a relative measure is used) and with respect to the margin of non-inferiority or the margins of equivalence. In the example the new treatment was noninferior. The figure can be used to show graphically the results of different analyses, eg, with or without adjustment (Figure 2) or ITT and per protocol.\(^6^n\)

Only 1 of 47 published equivalence or non-inferiority trials in ophthalmology evaluating prostaglandins depicted the CI graphically with the prespecified non-inferiority or equivalence margin.\(^2^n\)

**Discussion**

**Item 22: Interpretation.** Standard CONSORT item: Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence. Extension for non-inferiority trials: Interpret results in relation to the non-inferiority hypothesis. If a superi-
The present recommendations are among a series of extensions to the CONSORT Statement. The current versions of all CONSORT recommendations are available at http://www.consort-statement.org.

Author Contributions: Dr Piaggio had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Piaggio, Elbourne, Pocock, Evans, Altman.

Acquisition of data: Piaggio, Elbourne, Altman.

Analysis and interpretation of data: Piaggio, Elbourne, Altman.

Drafting of the manuscript: Piaggio, Elbourne, Pocock, Altman.

Critical revision of the manuscript for important intellectual content: Piaggio, Pocock, Evans, Altman.

Role of the Sponsors: The study sponsors had no role in the design and conduct of the study; the collection, management, analysis, and interpretation of the data; or the preparation, review, or approval of the manuscript.

Additional Contributions: We thank the members of the CONSORT Group for comments on earlier drafts. This group endorsed the submission for publication.

REFERENCES


8. Hopewell S, Clarke M, Moher D, et al; CONSORT Group. CONSORT for reporting randomized con-
REPORTING NONINFERIORITY AND EQUVALENCE

trolled trials in journal and conference abstracts: expla-
9. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for re-
11. Durkin ML, Palesch YY, pineau BC, Vining DJ, Cotton PB. The virtual colonoscopy study: a large mid-
testine clinic trial designed to compare two diagnost-
12. Clinical Outcomes of Surgical Therapy Study Group. A comparison of laparoscopically assisted and open col-
13. Chadwick D, vigabatrin European Monotherapy Study Group. Safety and efficacy of vigabatrin and car-
bamazepine in newly diagnosed epilepsy: a multicen-
ment of the Dose-Effect of a New Thrombolytic (ASSENT-2) Investigators. Single-bolus tenecteplase com-
15. Committee for Proprietary Medicinal Products. Note for Guidelines on Evaluation of Medicinal Products In-
dicated to Treat Bacterial Infections. London; England: European Medicines Agency (EMA); 2004.
17. Large and little trials are unethical because they disregard patients’ interests. Lancet. 2007;370(9602):1875-1877.
20. Moher D, Dubulcs CG, Wells GA. Statistical power, sample size, and their reporting in randomized con-
21. Garattini S, Berte V. Non-inferiority trials: are they supported by the evidence? Ann Intern Med. 2000;132(9):715-
722.
24. Eyawo O, Lee CW, Rachlis B, Mills EJ. Reporting of noninferiority and equivalence randomized trials for major prostaglandins: a systematic survey of the oph-
26. Krysan DJ, Kemper AR. Claims of equivalence in ran-
27. Dimick JB, diener-West M, lipsett PA. Negative re-
sults of randomized clinical trials published in the sur-
28. Costa LJ, Xavier AC, del Giglio A. Negative results in elderly patients: cardiovascular mortality and mor-
bidity the Swedish Trial in Old Patients with Hyper-
30. rothmann M, Li N, Chen G, Chi Gy, Temple R, Tsou HH. Design and analysis of non-inferiority mor-
31. Topol EJ, Moliterno DJ, Herrmann HC, et al; TARGET Investigators. Comparison of two platelet gly-
coprotein Ib/IIa inhibitors, ticlofiban and abciximab, for the prevention of ischemic events with percuta-
32. Warrier IK, Wang D, huong NT, et al. Can midlevel health-care providers administer early medi-
cal abortion as safely and effectively as doctors? a ran-
35. Zwarenstein M, Squire IB, Pogue J, Hart R, et al; ACTIVE Writ-
ing Group of the ACTIVE Investigators. Clopidogrel plus aspirin versus oral antiaggregation for atrial fibril-
lation in the Atrial fibrillation Clopidogrel Trial with Ibsen-
Haller for Prevention of Vascular Events (ACTIVE-
36. D’Agostino RB Sr, Massaro JM, Sullivan LA. Non-
inferiority trials: design concepts and issues—the en-
39. Willburger RE, Myers E, derby J, et al. Limuma-
coxib 400 mg once daily is comparable to indometha-
cin 50 mg three times daily for the treatment of acute flares of gout. Rheumatology (Oxford). 2007;46(7):
1126-1132.
40. Danis CW, McLaurin BT, Cox DA, et al; ACUTY Investigators. Bivalrudin for patients with acute coro-
41. Ferguson JI, Califf RM, Antman EM, et al; SYNERGY Investigators. Exenatide vs unfrac-
tioned heparin in high-risk patients with non-ST-
segment elevation acute coronary syndromes man-
42. giuliano AE, Hunt KK, Ballman KV, et al. Auxiliary dissection vs no auxiliary dissection in women with in-
vasive breast cancer and sentinel node metastasis: a ran-
45. Gunnell D, Saperia J, Ashby D. Selective serotonin reuptake inhibitors (SSRIs) and suicide in adults: meta-

©2012 American Medical Association. All rights reserved.