Supplementary Online Content


eAppendix.

This supplementary material has been provided by the authors to give readers additional information about their work.
eAppendix

Methodological considerations in noninferiority trials

**Hypotheses.** In a noninferiority clinical trial the aim is to establish that the effect of the new treatment, when compared to the reference treatment, is not below a pre-stated noninferiority margin $\Delta$ (alternative hypothesis). The null hypothesis is that its effect is above this margin. A trial may be planned to assess noninferiority for a primary endpoint and superiority for other endpoints typically related to safety.1

**Design.** Sample size depends on the level of confidence chosen, the desired power, the assumed percentages of success (binary outcomes) or the variability (continuous outcomes), and $\Delta$.2, 3 $\Delta$ can be specified as a difference in means or the logarithm of an odds ratio, risk ratio, or hazard ratio. It is often chosen as the smallest value that would be a clinically important effect.4 $\Delta$ should be particularly small to guard against the acceptance of inferior treatments. The required size of noninferiority trials is therefore usually larger than that for superiority trials because of the small size of the required margin.5

Sample size calculations are often based on the assumption that the difference in proportions or means between treatments is 0. [Ref 62] If instead we assume that the new treatment is less effective than the reference treatment, then the sample size has to be larger for a given power. Conversely, if the new treatment is more effective, the needed sample size for the same power is smaller.

**Conduct.** Trial conduct should closely match that of any trial that demonstrated efficacy of the reference treatment, provided they were of high quality.6 One should avoid features that might dilute true differences between treatments, thereby enhancing the risk of erroneously concluding noninferiority,7 such as poor adherence, dropouts, recruitment of patients unlikely to respond, and treatment crossovers. If there is an interest in showing superiority on another endpoint (or on the one for which demonstrating noninferiority is the main purpose of the study), then there is a clear incentive to conduct a trial that strives to separate the two treatments.

**Analysis.** Although a modified hypothesis testing framework exists,9, 10 a more informative CI approach is preferred in the design, analysis, and reporting of noninferiority and equivalence trials.11

For superiority trials, intention-to-treat (ITT) analysis (analyzing all patients within their randomized groups, regardless of whether they completed allocated treatment) is recommended as it is least biased analysis for assessing the effect of two policies.12 Full ITT analyses may not be possible when primary outcome data are missing, and when there are differences between policy and actual practice. ITT analysis often leads to smaller observed treatment effects than if analysis is limited to patients adhering to treatment (often called per-protocol (PP) analysis). PP analysis may be helpful as a sensitivity analysis. In noninferiority trials, however, ITT analysis may increase the risk of falsely claiming noninferiority,5 and the addition of the non-ITT analyses might be desirable as a protection from this risk.13 ITT analyses are not always anti-conservative in noninferiority trials,14 and alternative analyses that exclude patients not taking allocated treatment or otherwise not protocol-adherent could bias the trial in either direction.15 Whichever is chosen as primary, the other should be presented as a secondary sensitivity analysis. Given that both ITT and PP analyses may be problematic for type I error rates and bias in noninferiority trials, minimizing protocol deviations may be particularly important. In the presence of non-adherence, obtaining reliable data on its extent is important,16 and it has been suggested that a hybrid ITT/PP analysis might be preferable.17

Interim analyses in noninferiority trials have some differences in rationale from superiority trials. If noninferiority is established before the trial is completed, there may be no ethical requirement to stop early because of lack of efficacy.18 However, other advantages (adverse effects, cost) could justify stopping the trial, to expedite availability of the new treatment. If a treatment is clearly inferior, then stopping the trial (or a particular trial arm) is ethically justified.19-23 Stopping rules might be asymmetric, a trial being allowed to continue longer if the new treatment appears superior,23 although this result is unlikely.18 Confidence intervals for treatment effects should be adjusted for interim analyses, much as the criteria for statistical significance should be adjusted to control type I error rates.

**Interpretation.** Interpreting a non-inferiority trial’s results depends on where the CI for the treatment effect lies relative to both the margin of noninferiority $\Delta$ and a null effect. The upper bound of the 2-sided $(1-2\alpha)\times100\%$ CI or that of the 1-sided $(1-\alpha)\times100\%$ CI for the treatment effect has to be below the margin $\Delta$ to declare that noninferiority has been shown, with a significance level $\alpha$. Both $\Delta$ and $\alpha$ should be prespecified in the noninferiority hypothesis.
Many noninferiority trials based their interpretation on the upper limit of a 1-sided 97.5% CI, which is the same as the upper limit of a 2-sided 95% CI. Although both 1-sided and 2-sided CIs allow for inferences about noninferiority, we suggest that 2-sided CIs are appropriate in most noninferiority trials.\textsuperscript{23} If a 1-sided 5% significance level is deemed acceptable for the noninferiority hypothesis test, a 90% 2-sided CI could then be used. Figure 1 interprets several possible scenarios with 2-sided CIs for a noninferiority trial.

Once noninferiority is evident, it is acceptable to then assess whether the new treatment appears superior to the reference treatment, using an appropriate test or CI, with a significance level defined a priori and with an ITT analysis.\textsuperscript{24,25}

It is inappropriate to claim noninferiority post hoc from a superiority trial unless the noninferiority hypothesis was stated in the protocol and clearly related to a predefined margin of equivalence. That is, both superiority and noninferiority hypotheses need explicit specification in the trial protocol and appropriate methods applied for sequential testing.\textsuperscript{26,27} It is, however, always reasonable to interpret a CI as excluding an effect of a particular prestated size.\textsuperscript{28}

**Choice of the noninferiority margin $\Delta$**

The choice of $\Delta$ should be justified on statistical and clinical grounds. The noninferiority margin $\Delta$ has been given different names: irrelevant difference\textsuperscript{29}, clinical acceptable amount\textsuperscript{30}. It is often determined to preserve a pre-specified fraction of the reference treatment effect relative to placebo. However, if the reference treatment has a large effect over placebo it does not mean that large differences are unimportant, therefore it might be unwise to decide to preserve a fixed fraction without considering the resulting margin with a clinical perspective.\textsuperscript{131} The choice of $\Delta$ needs to take into account the uncertainty of the difference between the reference treatment and placebo. However, there are situations where no such past placebo-controlled evidence exists, e.g., in device or surgical trials or where the reference treatment is so long established that no relevant data exist. In that case, other judgements come into play as to what is a sensible choice of $\Delta$.

The FDA guidance\textsuperscript{32} about the design of noninferiority trials are generally conservative in their use of these trials to provide sufficient evidence to grant a license. They therefore look to narrow margins for the noninferiority and good evidence of “assay validity”. The EMA might be seen as slightly more liberal, and their “Points to Consider” on switching from Superiority to Noninferiority (2000) and their Guidance on choice of NI margin (2005)\textsuperscript{12} concentrate on these issues. The EMA emphasizes confidence intervals more than hypothesis testing in this context, but also give importance to assay validity.

If past placebo-controlled trials exist establishing the efficacy of the reference treatment, then there are two steps in the determination of $\Delta$:

1. Estimate the difference between the reference treatment and placebo. Given several previous trials, the effect of the reference treatment can be estimated from a meta-analysis. For the fixed margin approach the Federal Drug Administration of the United States (FDA) suggests taking the lower bound of the 95% confidence interval of the effect size as the estimate\textsuperscript{32}. In doing this, it is assumed that there is assay sensitivity, e.g., assurance that the reference treatment would be superior to a placebo if a placebo is employed in the current trial; and that the constancy assumption holds (e.g., that the difference between the reference treatment and placebo in those trials will hold in the setting of the new trial if a placebo control is used). In many circumstances this is too stringent a criterion, leading to an unrealistically small $\Delta$, e.g., in the PROFESS trial.\textsuperscript{33}

2. Once the effect of the reference treatment is estimated, $\Delta$ should be determined using clinical judgment so that at least a certain amount of the superiority of the reference treatment over placebo is retained\textsuperscript{32} (e.g., 50% to 80% depending on the seriousness of the outcome).

It needs to be recognized that such statistical logic can lead to an unachievable sample size. Hence choice of $\Delta$ needs to take account of such practicality.

**Approaches to analysis of a noninferiority trial if past placebo-controlled trials exist**

1) Fixed margin or two confidence interval method: Conclude noninferiority if the upper bound of the 95% confidence interval for the difference between the reference treatment and the new treatment (for a beneficial outcome) is below $\Delta$ in the NI trial. This method is called ‘two confidence interval’ because it assumes that $\Delta$ has been determined using the lower bound of the 95% confidence interval to estimate the effect of the reference treatment as described above.\textsuperscript{32}

2) Synthesis method: combines the estimate of the effect in the NI trial with the estimate of the reference treatment effect from a meta-analysis of historical trials, making use of the variability from both. With this method, $\Delta$ is not predetermined. It is controversial, since it gives an estimate of the difference between the new treatment and placebo using indirect evidence. It is not usually recommended as a primary analysis.\textsuperscript{32}

© 2012 American Medical Association. All rights reserved.
Noninferiority and effectiveness of the new treatment

To demonstrate noninferiority and effectiveness of the new treatment compared to the reference treatment, the following are needed:

1. Noninferiority margin was defined a priori on the basis of statistical reasoning and clinical judgment; the trial was planned with adequate power and appropriate significance level.

2. The effectiveness of the reference treatment has been established, assay sensitivity and constancy assumptions are valid.

3. The new treatment should be shown to be beneficial over the reference treatment regarding at least one non-eficacy outcomes, such as safety, tolerability, cost, or convenience. This is the basis for the rationale of conducting a noninferiority trial. If the other benefits of the new treatment have not been shown in the past, there should be power to show them in the current noninferiority trial, using a superiority hypothesis.

4. An attempt to demonstrate superiority of the new treatment to a putative placebo might be required by some regulatory authorities if at least one placebo-reference treatment trial has been conducted in the past. It is calculated by combining evidence from a) the difference between the new and the reference treatment in the present trial and, b) the difference between the reference treatment and placebo in historical trials. This method is subjected to all the caveats of indirect comparisons and the potential bias induced by any lack of "assay constancy".

More detailed lists of essential elements to assess noninferiority have been provided.

Recent research in methodology

Determination of the margin of noninferiority

Methods to determine the margin of noninferiority for a mean difference of a continuous measurement have been discussed and extended to the ratio of two means or the hazard ratio.

To address the caveats involved in the determination of a fixed $\Delta$, methods have been developed for the determination of the margin of noninferiority based on adaptive testing, for both continuous and binary outcomes, that allow it to vary as a function of the reference treatment response. This strategy is useful to assure that the new treatment has no adverse impact on a main outcome (e.g. in oncology trials, assuring that 'survival of patients administered the new therapy (is) at least 95% of survival with standard care'), while offering advantages in safety or other clinical outcomes. It also has the advantage of being a more efficient testing procedure, thus saving on sample size.

The situation in which the reference treatment response cannot be predicted accurately presents difficulties. Tests of noninferiority have been developed for the situation in which the noninferiority margin is a function of unknown parameters.

Estimation of the effectiveness of the new treatment in relation to placebo

In recent years it has become clear that it is sometimes not sufficient to demonstrate noninferiority of the new treatment compared to the reference treatment, but it is also required to demonstrate the effectiveness of the new treatment (see section above about Demonstrating noninferiority and effectiveness of the new treatment). Unlike superiority trials, interpretation of noninferiority trials can depend on information that is not measured in the trial, namely that the reference treatment is effective compared to placebo. If this evidence is not available, a noninferiority trial might be showing noninferiority of a new treatment with respect to an ineffective reference treatment. This point has been raised mainly in the context of drug development, but it is relevant also for other treatments or interventions.

A unified approach for establishing the efficacy of a new treatment, with the fixed margin approach and the synthesis approach being particular cases, has been proposed. This approach allows to control the impact of departures from the assumptions of assay sensitivity and constancy.

Analysis

A method of analysis that preserves the integrity of randomization more effectively than a per-protocol analysis has been developed for time-to-event outcomes.

Different methods to calculate confidence intervals exist and they could result in different conclusions regarding the noninferiority hypothesis.
eReferences


