Review article
http://dx.doi.org/10.3345/kjp.2012.55.11.403
Korean J Pediatr 2012;55(11):403-407
eISSN 1738-1061 • pISSN 2092-7258

Understanding noninferiority trials

Seokyung Hahn, PhD

Department of Medicine, Seoul National University College of Medicine, Seoul, Division of Medical Statistics, Medical Research Collaborating Center, Seoul National University Hospital, Seoul, Korea

Received: 29 June, 2012, Accepted: 18 August, 2012

Corresponding author: Seokyung Hahn, PhD
Department of Medicine, Seoul National University College of Medicine, 103 Daehak-ro, Jongno-gu, Seoul 110-799, Korea
Tel: +82-2-740-8911, Fax: +82-2-743-8361
E-mail: hahns@snu.ac.kr

Noninferiority trials test whether a new experimental treatment is not unacceptably less efficacious than an active control treatment already in use. With continuous improvements in health technologies, standard care, and clinical outcomes, the incremental benefits of newly developed treatments may be only marginal over existing treatments. Sometimes assigning patients to a placebo is unethical. In such circumstances, there has been increasing emphasis on the use of noninferiority trial designs. Noninferiority trials are more complex to design, conduct, and interpret than typical superiority trials. This paper reviews the concept of noninferiority trials and discusses some important issues related to them.

Key words: Noninferiority trial, Controlled clinical trials, Randomized controlled clinical trials, Clinical trials

Background

Treatment efficacy is considered as the capacity of a given intervention to produce a beneficial effect. The use of control groups in clinical trials is important for differentiating patient outcomes caused by experimental treatments from those caused by other factors, such as natural disease progression. Efficacy of an experimental treatment is most convincingly established by demonstrating its superiority to a placebo in a placebo-controlled trial, by showing superiority to an active control treatment, or by demonstrating a dose-response relationship. This type of trial is referred to as a superiority trial[1,2].

The term "active control trial" refers to clinical trials in which the control treatment employed is an active one. There are several reasons for using active controls in clinical trials[1,3,4]. For example, in trials involving serious outcomes such as mortality, it is unethical to use a placebo when active treatments are available. Clinical equipoise, referring to the state of true uncertainty about the relative benefits of alternative treatments under the "null" hypothesis to be tested, is an ethically necessary condition in all clinical research[5]. Active controls are sometimes used to demonstrate the efficacy of a drug that may have large placebo effects. Active controls are also used to determine how experimental treatments compare to alternative treatments. Active control trials aim to demonstrate that treatments of interest have either superior effects or similar effects to the control.

Our interest usually lies in being able to demonstrate that a particular new treatment can be recommended as being better than existing treatments. Such trials are known as superiority trials, where we seek sufficient evidence to reject the hypothesis that "the 2 treatments have equal effects" in favor of the superiority of the new treatment. However, failure to observe sufficient evidence for rejection of the null hypothesis does not necessarily suggest the equivalence of 2 treatments[2].

If the intent of a study is to demonstrate that differences between control and experimental treatments are not large in either direction, then it is known as an equivalence trial. Bioequivalence trials are those in which generic drug preparations are compared to currently marketed formulations with respect to pharmacokinetic parameters in order to evaluate their *in vivo* biological equivalence, and those are the ones where showing equivalence between treatments is truly of interest. When determining the effects of experimental treatments on clinical end points, it would not be sensible to investigate whether their effects are no worse than, as well as no better than, those of the control. Although noninferiority and equivalence trials have often been both referred to as "equivalence trials," they are distinct. If the intent of a study is to demonstrate that an experimental treatment is not substantially worse than a control treatment, the study is known as a noninferiority trial. However, there are some complicated issues with trials of this type that make them less reliable than typical superiority trials[6-9]. Some guidelines have been provided by regulatory bodies[10-12], and the CONSORT (Consolidated Standards of Reporting Trials) Group has also published an extension of their guidelines for such trials[13].

Demonstrating treatment effect

When trying to demonstrate that a new treatment is better than a placebo or active control, we statistically analyze the trial data to determine whether the result provides sufficient evidence to reject the null hypothesis, i.e., that the 2 treatments have the same effect. For example, if in a randomized controlled trial a reduction in blood pressure of 15 mmHg is observed for a newly developed antihypertensive drug, while an active control induces a reduction of only 5 mmHg reduction, it would be incorrect to conclude that the new drug is more efficacious based on the observed difference of 10 mmHg because such a difference could be due to chance of sampling. Instead, we calculate a confidence interval (CI) around the observed difference, which allows a degree of uncertainty associated with the observed value of 10 mmHg, and it is within this CI that the true difference will likely lie. For example, if a 95% CI is calculated (3, 17), this means that when samples are repeatedly drawn and treatment differences are estimated, 95% of the estimated differences would be expected to be a value between 3 and 17, and one of those could be the true difference. Some values may also fall outside the CI and although these could also be the true value, this likelihood is very small. Therefore, in this example, we conclude that the true difference is very unlikely to be 0, and there is sufficient evidence to conclude that the new treatment is superior to the control when the *P* value is <0.05 (Fig. 1). Note, that if due to a small sample size the CI is very wide, then the study is not likely able to demonstrate superiority. Thus, concluding equivalence or noninferiority on the basis of a nonsignificant test of the null hypothesis, i.e., no difference between the experimental treatment and the active control is inappropriate.

Why noninferiority trials?

Noninferiority trials may be performed to demonstrate that a new treatment is better than an assumed placebo in situations where conducting a placebo control trial is unethical. They may also be used when the new treatment may offer important advantages over currently available standard treatments, in terms of improved safety, convenience, better compliance, or cost. In addition, clinical trials are increasingly required to demonstrate benefits in clinical endpoints rather than surrogate endpoints, even though the incremental benefits from new treatments is diminishing, which is also an important factor in determining sample size. Such practical considerations are also driving a trend towards designing clinical trials that aim to demonstrate experimental treatments have similar effects to active controls of a proven efficacy rather than a superior effect. However, testing for noninferiority makes trial design and interpretation of results less straightforward than typical superiority trials.

Analysis of noninferiority trials

A naïve approach to analyzing experimental data from active control trials is to compare new and control treatments in the standard way and, if no difference is detected, to declare the treatments equivalent. However, a problem with this approach is that if the sample size is too small, so that the CI is too wide, equivalence could be inappropriately concluded. Thus, if the aim is to assess true "equivalence," then the null hypothesis should be that the treatments differ, and whether there is sufficient evidence to declare their equivalence and reject the null hypothesis should be investigated. Since it is practically impossible for 2 treatments to have exactly equivalent effects, "equivalence" in terms of clinical evaluation means that the effects of 2 treatments differ by no more than a tolerable amount, known as the equivalence margin. In an equivalence trial, if the effects of the 2 treatments differ by more than

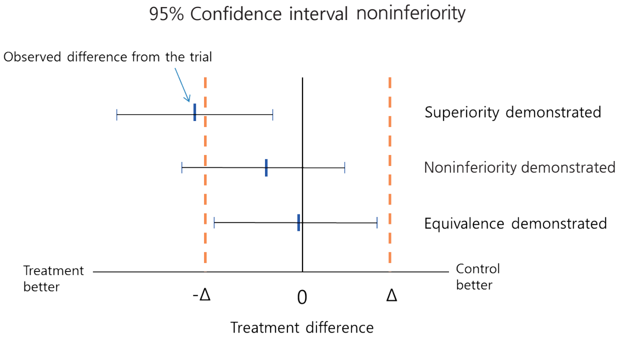

Fig. 1. Testing superiority, equivalence/noninferiority. Δ: margin for equivalence/noninferiority.

the equivalence margin in either direction, then equivalence does not hold. Noninferiority trials, on the other hand, aim to show that an experimental treatment is not less effective than an active control by more than the equivalence margin. In a trial intending to show that there is a difference less than a specific amount between control and experimental treatments, a noninferiority design statistically tests the null hypothesis that the experimental treatment is inferior by the equivalence margin. For example, in a trial where an outcome of higher values is desirable, if the upper limit of the 95% CI for the treatment difference by the experimental treatment is less than the equivalence margin, then the null hypothesis of inferiority can be rejected at the 5% significance level (Fig. 1).

Determination of noninferiority margins

The margin (Δ), the maximum acceptable extent of clinical noninferiority of an experimental treatment, must be prospectively defined. One approach to specifying the margin is based on clinical significance, which can obviously be subjective. Sometimes it is possible to choose a margin for declaring noninferiority of a treatment, in which that treatment ends up having no effect or even a detrimental effect. For example, let us assume that it is known from the literature that a treatment response to a control drug is somewhere between 15% and 30%. If the control drug has a response less than 20% and the margin was set at 20%, we could conclude that the new treatment is noninferior, even if it exerts no response. Such a scenario could be possible because the lower limit of the range for the control treatment response is 15%. The margin should be based on both statistical reasoning and clinical judgment and, in the setting of a placebo-controlled trial, cannot be greater than the smallest response that could be reliably expected from the active treatment compared to a placebo. Using a treatment difference for the control treatment based on a previously published placebo-controlled trial, would be a way to consider the size of the margin statistically. If we let T, C, and P represent the efficacy values of a new treatment, an active control, and a placebo, respectively, then in a trial where higher efficacy values are desirable, the standard null and alternative hypotheses for proving noninferiority are; H0: $C–T≥Δ$ (T is inferior to C) and H1: $C–T<Δ$ (T is noninferior to C), respectively. The alternative hypothesis (H1) states that the new treatment may have a negative effect compared to the active control, but by no more than Δ. For a new treatment to at least have better efficacy than the placebo while it can be less effective than the active control within the extent of Δ, the size of the margin allowed to the maximum limit would be the entire effect size of the control treatment. A demonstration that the difference between a control and an experimental treatment is less than Δ, would indicate that the new treatment is better than the placebo, i.e., it is effective. If there were a need to ensure this conclusion, the margin could be chosen to be a fraction of the control treatment effect (Fig. 2). For example, the margin could be 50% or 25% of the entire control treatment effect compared to placebo.

Does noninferiority imply a treatment is effective?

As previously pointed out, setting an inappropriate margin can cause a noninferiority test to misleadingly conclude a ineffective treatment to be effective. In some cases, noninferiority tests can be useless, unless the trial is carefully designed. A clinical trial should have the ability to distinguish effective treatments from those that are less effective, or ineffective. This is defined as "assay sensitivity" and there is a question of whether noninferiority trials have the power to detect a beneficial treatment against a placebo even if a placebo group is included in the trial. For example, even if a control treatment has shown efficacy in previous placebo-controlled trials, unless one can reliably expect that the control treatment effect consistently occurs in the current trial and if both treatments are truly ineffective, the test may just declare the noninferiority of an experimental treatment to an ineffective control (Fig. 3).

The presence of assay sensitivity in a noninferiority trial is not verifiable but may only be assumed based on historical evidence of

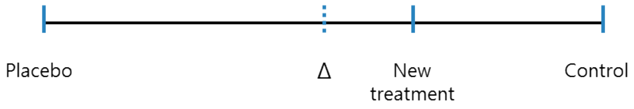

Fig. 2. Noninferiority margin. The positioning of the outcome result for each treatment is indicated. Δ: margin for equivalence/noninferiority.

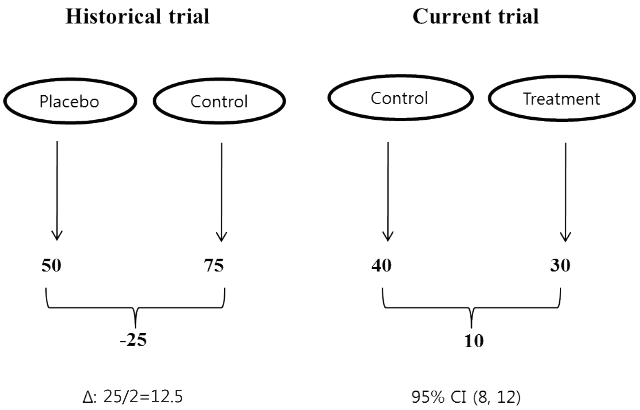

Fig. 3. Hypothetical scenario. Numbers represent the values of a positive outcome. The noninferiority margin is determined by halving the control effect, based on a historical placebo-controlled trial. The upper limit of the 95% confidence interval (CI) for the treatment difference by the new treatment compared to the control is less than the margin to conclude noninferiority, although there is an implication of incomparability. Δ: margin for equivalence/noninferiority.

sensitivity to drug effects, the similarity of trial designs to those that were able to distinguish efficacy of the active control from that of a placebo, and the quality of trial conduct. Trial designs should be compared closely in terms of inclusion criteria, methods of diagnosis, and concomitant treatments used to evaluate consistency over time. The notion of assessing noninferiority in this context is similar to an indirect comparison, i.e., we indirectly evaluate the superiority of an experimental treatment over a placebo using a hypothetical treatment difference for the experimental treatment compared to the placebo that is indirectly measured based on the treatment differences between the active control and the experimental treatment and between the control and the placebo.

Sample size considerations

The sample size of a noninferiority trial is very sensitive to the expected effects of the experimental treatments and controls. Although there could be other reasons for undertaking noninferiority trials, showing noninferiority would be more appropriate when there is an expectation that 2 treatments are similar. A larger sample size is needed if a new treatment is assumed to be slightly less effective than the control, since in such situations it is more difficult to show noninferiority, unless a considerably narrower CI is obtained. On the other hand, the required sample size can be reduced if a new treatment is assumed to be slightly more effective than the active control. The noninferiority margin is another major factor that influences sample size, and the greater the tolerance that is allowed, the smaller the sample size that is needed. However, an inflated margin may cause considerable loss of statistical power if noninferiority can be accepted only by a smaller margin (Table 1).

Choosing the analysis population

Intention-to-treat (ITT) is conventionally accepted as an unbiased analytical approach for superiority trials. Analysis of all randomized patients, according to the treatments to which they were assigned, regardless of whether they received the treatment or not, confers a conservative effect on the outcome of the trial. However, ITT analysis may not be conservative for noninferiority trials, since including dropouts in the analysis tends to bias the results toward equivalence, even when an experimental treatment is less effective than the control. The per-protocol analysis, which includes all patients who satisfactorily complied with the assigned treatment and who had no major protocol violations, is more likely to identify any treatment differences, but it can also substantially bias the results in either direction. The recommended approach for noninferiority trials is to perform both analyses and to conclude noninferiority if both analysis produce the same result.

Table 1. Sample Size in Noninferiority Trial

Control (%)	Experimental (%)	Δ (%)	Power (%)	n/group
60	60	5	80	1,188
60	22	5	80	601
62	60	5	80	3,268
60	60	8	80	464
60	60	5	46	464

Numbers represent the values of a positive outcome (e.g., response rates).
Δ: margin for noninferiority.

Switching between superiority and noninferiority

Interpreting a noninferiority trial as a superiority trial is credible and without a need for a statistical penalty for multiple testing. If the 95% CI for the treatment benefit excludes not only the noninferiority margin but also zero, it would be considered adequate evidence to prove superiority within the same trial. However, the opposite approach is not valid. If a superiority trial fails to reject the null hypothesis but the trial data appear to suggest treatment equivalence, one may also be tempted to infer noninferiority. If there is a possibility for testing noninferiority alongside a superiority test, one should predefine both hypotheses with a justifiable margin for noninferiority in the protocol. Testing noninferiority based on an ad hoc determination of a noninferiority margin after a trial is complete would not be acceptable due to bias. When both hypotheses are carefully planned within a protocol, both can be tested on the same population without a statistical penalty.

Conclusions

With improvement in health technologies, standard care, and clinical outcomes, the incremental benefits of newly developed treatments may only be marginal over existing treatments. Sometimes assigning patients to a placebo is unethical, and in such circumstances, there is increasing emphasis on the use of noninferiority trial designs. Noninferiority trials are more complex to design, conduct, and interpret than conventional superiority trials. When planning a noninferiority trial, one should adequately understand its concept and the possible drawbacks. Choice of the noninferiority margin is critical in designing noninferiority trials. One reasonable way to define a margin is to base it on some proportional effect that the active control has shown over placebo in previous studies. However, the margin should be based on both statistical reasoning and clinical judgment. A justifiable margin for noninferiority should be predefined in the protocol. It should reflect uncertainties in the evidence on which the choice is based and should also be suitably conservative. Testing noninferiority based on *ad hoc* determination of the margin after a trial is complete is not acceptable. Noninferiority trials are based on some directly unverifiable assumptions. In order to demonstrate assay sensitivity, it is important to evaluate the conditions of previous trials as closely as possible, including trial design and patient characteristics. Therefore, when planning a noninferiority trial, all necessary considerations should be taken to ensure that false claims of noninferiority are avoided.

References

1. Rosato R, Ciccone G, Bo S, Pagano GF, Merletti F, Gregori D. Evaluating cardiovascular mortality in type 2 diabetes patients: an analysis based on competing risks Markov chains and additive regression models. J Eval Clin Pract 2007;13:422-8.
2. Stephan BC, Kurth T, Matthews FE, Brayne C, Dufouil C. Dementia risk prediction in the population: are screening models accurate? Nat Rev Neurol 2010;6:318-26.
3. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. Ann Intern Med 2000;133:455-63.
4. Ellenberg SS, Temple R. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 2: practical issues and specific cases. Ann Intern Med 2000;133:464-70.
5. Freedman B. Equipoise and the ethics of clinical research. N Engl J Med 1987;317:141-5.
6. D'Agostino RB Sr, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. Stat Med 2003;22:169-86.
7. Gotzsche PC. Lessons from and cautions about noninferiority and equivalence randomized trials. JAMA 2006;295:1172-4.
8. Jones B, Jarvis P, Lewis JA, Ebbutt AF. Trials to assess equivalence: the importance of rigorous methods. BMJ 1996;313:36-9.
9. Pocock SJ. The pros and cons of noninferiority trials. Fundam Clin Pharmacol 2003;17:483-90.
10. European Medicines Agency. Committee for proprietary medicinal products (CPMP) [Internet]. London: European Medicines Agency; c2012 [cited 2012 Jun 5]. Available from: http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003658.pdf.
11. European Medicines Agency. Guideline on the choice of the non-inferiority margin [Internet]. London: European Medicines Agency; c2012 [cited 2012 Jun 5]. Available from: http://www.emea.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf.
12. U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER) Guidance for Industry, non-inferiority clinical trials [Internet]. Washington: U.S. Department of Health and Human Services; c2012 [cited 2012 Jun 5]. Available from: http://www.hhs.gov/asl/testify/2012/04/t20120418a.html.
13. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. JAMA 2006;295:1152-60.