Review Article

Non-inferiority trials and non-inferiority margin: an overview

Kourosh Zali

Arya Research Institute, Isfahan, Iran

*. Corresponding author: Kourosh Zali, Arya Research Institute, Iran, E-mail: k.zali.arya.res.inst@gmail.com.

Abstract

The randomized controlled trial (RCT) is considered as the best interventional design to assess issues related to treatment and prevention. The RCTs can have different designs including superiority, equivalence or non-inferiority design. A superiority trial aims to detect the potential superiority of a new therapy compared to an active comparator or a placebo, an equivalence trial tends to demonstrate that a new therapy is equivalent (within margins) to its active comparator, and non-inferiority trial (NIT) is going to show that the new therapy is not worse than the comparator, as a typical active drug. Increasingly, major trials are conducted to see if the efficacy of a new treatment is as good as a standard treatment. The new treatment usually has some other advantages (e.g., fewer side effects, ease of administration, lower cost), making it worthwhile to demonstrate non-inferiority in respect to efficacy. Thus, NIT is going to determine whether a new treatment is not worse than a reference treatment by more than an acceptable amount. Among the challenges of NITs compared with superiority trials are the choices of the non-inferiority margin (NIM), the primary population for analysis, and the comparator treatment considering several choices for the comparator arm in a NIT. This article is going to review the current knowledge about NIM.

Keywords: Randomized controlled trials, Drug development, Superiority, Equivalence, Non-inferiority trial, Non-inferiority margin

Background

The randomized controlled trial (RCT) is currently considered as the best experimental design to assess issues related to treatment and prevention (1). Classically, the defined medical experiments are projected to determine which of two or more interventions is the most effective after the randomized allocation of patients to different study groups. One of the groups is considered as the control, sometimes may refer to the absence of treatment, placebo or, more often, a treatment of recognized efficacy. These models are called superiority trials, whose objective is to determine whether a treatment under investigation is superior to the comparative drug or not (2).

The presence of a placebo arm is certainly a substantial element in control trials, because the effectiveness of efficacy of a new therapy can be evaluated through a direct comparison between the test treatment and the placebo arm (3). Moreover, the placebo controlled RCT, as a gold standard for determining the efficacy of a new therapy (4, 5), is trying to demonstrate the superior efficacy (superiority) of the new drug or treatment over placebo (4). However, the conducting of a placebo controlled RCT is frequently very difficult or even impossible. To solve this problem, the experimental therapy can be compared to an established treatment, referred to a comparison group, instead of placebo (4). In fact, in recent decades, the availability of standard treatments and ethical concerns have led scientists to consider an active or positive control treatment as a comparator to assess the treatment effect without a placebo arm. Such an assessment is often made under a so-called "non-inferiority" trial design (3).

Different designs of RCTs

The RCTs can have different designs, in the process of drug development, including superiority, equivalence or non-inferiority design. A superiority trial aims to detect the potential superiority of a new therapy compared to an active comparator or a placebo, an equivalence trial tends to demonstrate that a new therapy is equivalent (within margins) to its active comparator and non-inferiority trial (NIT) is going to show that the new therapy is not worse than the comparator, as a typical active drug (6, 7).

It should be noted that NIT and equivalence trials are sometimes, mistakenly, used interchangeably (8). Moreover, assessing the frequency of NIT is complicated, because not all NIT or equivalence trials use these words, and the term "equivalence" is often inappropriately used when reporting "negative" (null) results of superiority trials (9). NITs, comparing a new treatment with a standard, are indeed becoming frequent because of the need to replace standard treatments by other treatments having comparable efficacy but presenting other advantages (9).

There is a clear rationale for the classification of superiority/non-inferiority in the regulatory trials (10). An important reason for conducting a NIT is when a new therapy is expected to have advantages over the standard therapy, other than the main therapeutic effect (11). In addition, NITs are performed when the main therapeutic effect of the new therapy is expected to be not unacceptably worse than that of the standard therapy, and the new therapy is expected to have advantages over the standard therapy in costs or other (health) consequences (12). In recent decades, active controlled trials are indeed often performed instead of or in addition to placebo-controlled trials as the basis for marketing authorization and reimbursement decisions (5). It is also widely accepted that there are important differences between superiority and non-inferiority trials in terms of their design, analysis and interpretation (10). Furthermore, in RCT, using placebo is unethical for treating the patients who have critical, severe or life-threatening diseases such as cancer when approved and effective therapies such as standard treatment or active control drugs exist (13).

Non-inferiority randomized controlled trials

The NITs have great applicability in oncology, preventive cardiology and in the assessment of anti-infective agents (2). The regulatory authorities have also been requiring non-inferiority trials for the assessment of biosimilars (products obtained by biotechnological processes such as therapeutic proteins, monoclonal antibodies and so on). A product that has proved to be non-inferior in relation to an established treatment regarding an efficacy variable,

however, may present important advantages such as better tolerability, use convenience, advantages, different metabolic pathways, and less interactions (2). The concept of a NIT design (instead of superiority design) was introduced as a new study to assess whether a new therapy has efficacy that would be similar to, or at least not much worse than (non-inferior to), a standard therapy (14). In addition, NITs were originally developed in the setting of drug approval, where regulatory agencies have to make a binary decision, to give the license to the new treatment or not (10).

Increasingly, major trials are conducted to see if the efficacy of a new treatment is as good as a standard treatment (15). The new treatment usually has some other advantages (e.g., fewer side effects, ease of administration, lower cost), making it worthwhile to demonstrate non-inferiority in respect to efficacy (16). NIT comparing a new treatment with an active control drug or a standard treatment is often preferred. The main goal of a NIT would indeed be the assessment of non-inferiority of the new treatment by demonstrating that the new treatment is not inferior to (or at least is as effective as) the active control drug or the standard treatment (13). Thus, NIT is going to determine whether a new treatment is not worse than a reference treatment by more than an acceptable amount (9). The objective of a NIT is to demonstrate that the intervention being evaluated achieves the efficacy of the established therapy within a predetermined acceptable NIM (14).

A NIT is reasonable when a new treatment has some property sufficiently favorable that physicians, and their patients, would be willing to sacrifice some degree of benefit relative to an already approved therapy (11). The advantage could be reduced cost, improved ease of use or dosing schedule (monthly versus weekly injections), simpler storage (not requiring refrigeration), or an improved safety profile. The benefit given up in exchange for these advantages, however, should not be so large that patients and physicians are not willing to use the new product (11). In addition, US Food and Drug Administration (FDA) guidelines state that non-inferiority design for a clinical trial is chosen when it would not be ethical to use a placebo, or a no treatment control, or a very low dose of an active drug, because there is an effective treatment that provides an important benefit (e.g., life-saving or preventing irreversible injury) (17).

Challenges of Non-inferiority trials

Due to different nature of the superiority and NIT designs (18), in the analysis and interpretation of

NITs, at least five factors must be carefully considered to ensure the validity of the study: selection of NIM, number of patients needed for the study, control of study sensitivity, definition of population analysis and ethical justification (19).

In a NIT, the sample size calculation is conventionally based on achieving adequate power to demonstrate that the relevant confidence limit excludes the specified NIM, assuming that the two treatments are equally effective (10). The magnitude of an NIM critically determines the size of a trial (i.e., trial size increases inversely to the square of the margin) and is of foremost importance when interpreting its results (20). Different NIMs may lead to very different sample sizes required for achieving a desired power for establishing non-inferiority of the test treatment. NIMs may be selected based on the previous experiences in placebo control trials under similar conditions to the new trial. The significance level of 2.5% should be used when performing a one-sided non-inferiority testing. A narrower margin requires a much larger sample size for achieving the desired power for establishing non-inferiority (13).

The method of choice for analysis of non-inferiority trials consists in the construction of confidence intervals (Cis), usually 95% CI. The treatment is considered non-inferior if the inferior limit of the 95% CI of the difference between treatment and control does not include the value of the specified margin (2). Non-inferiority of the new therapy would be demonstrated if the lower confidence limit for the difference in effect between the therapies turns out to lie above NIM; thus, the NIT is designed as a one-sided trial (21). The recommended approach by regulators, such as the US FDA, is to compare the estimated 95% CI of the new drug vs. the active comparator from the NIT to a predefined NIM. If the CI lies entirely below the margin, non-inferiority of the new drug to the active comparator can be concluded. This then demonstrates that even if the difference is in the favor of the active comparator, it does not exceed the unacceptably worse criteria of non-inferiority (NIM) (22). Non-inferiority is demonstrated if the lower confidence limit lies above or to the right of the NIM (23).

Non-inferiority margin (NIM)

In practice, one of the key issues in a NIT is the selection of an appropriate NIM (13). The NIM is based on the proportion of therapeutic effect of the active control that should be retained (24). In a NIT without a placebo arm, a basic and vital design specification is the NIM, which is required to be smaller than or no greater than the effect of the selected active control under the NIT setting (3).

Equivalence margins are often far too large to be clinically meaningful and that a claim of equivalence may be misleading if a trial has not been conducted to an appropriately high standard (21). In fact, the most fundamental design specification of a NIT is the so-called NIM. The FDA recommends that, as a minimum requirement, the NIM must be selected to enable the trial to demonstrate that had a placebo been present in the trial, the test treatment would have been more effective than the placebo (3). The NIM— the value that allows for a new treatment to be 'acceptably worse'—is used as a reference for conclusions about non-inferiority. It is recommended that NIM is chosen on a clinical basis, meaning the maximum clinically acceptable extent to which a new drug can be less effective than the standard of care and still show evidence of an effect (8).

The NIM, is a critical component when considering the definition of "not being worse" in NITs. This non-inferiority parameter defines the boundary not to be exceeded by the upper confidence limit of the difference between study treatments' event rates; measured in absolute percentages or ratios; i.e., relative risk (RR), odds ratio (OR), or hazard ratio (HR). The NIM is fixed in advance and should be clinically justified. It ideally represents the smallest evidence of inferiority that, if true, would mean the new treatment is unacceptable (20).

Since no placebo arm exists in such a trial, many difficult issues arise, such as choice of the NIM, proper statistical method of non-inferiority testing, and level of statistical evidence (3). There are several design considerations that are unique to NITs that need to be considered. The first, and possibly most important, is the determination of the NIM. The NIM is the degree to which the new therapy can be less efficacious than the established treatment and still be considered non-inferior (25). The analysis of non-inferiority depends on the NIM that is the largest clinically acceptable difference between the new drug and the active comparator (22).

The NIM is used to assess whether the test drug will preserve what is considered a clinically significant fraction of the effect of the active comparator. To that aim, historical evidence on the active comparator from (a meta-analysis of) placebo-controlled (and/or active controlled trials) is used (26).

Determining NIM

NITs present some methodological challenges, especially in determining the NIM (5). Regulators recommend that the NIM should be defined based on statistical considerations and clinical judgement (22). The magnitude of NIM depends on what would be a clinically important difference, the

expected event rates and regulatory requirements, the known effect of the standard treatment vs placebo, the severity of the disease, toxicity, inconvenience, cost of the standard treatment, and the primary endpoint (15). A smaller non-inferiority margin is likely appropriate if the desired disease is severe or if the primary end point is death (14).

The NIM, the maximum acceptable extent of clinical non-inferiority of an experimental treatment, must be prospectively defined (27). 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 non-inferior, 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% (27).

Methods of determination of NIM

There is no well-established method to determine the NIM, it is very important that this margin be pre-specified and the criteria for how it was established well defined prior to conducting the study (4). Defining the NIM is crucial, yet one of the most challenging aspects in the design of non-inferiority trials (28). There are different methods to determine NIM including:

1- Some studies have considered a fixed NIM of 12% or around 12% (19).

2- Usually, 50 to 75% is accepted as the fraction of the estimated control effect to be preserved in relation to the placebo. To preserve 50%, NIM must be equal to 10% and, to preserve 75%, NIM must be equal to 5%. At this point, clinical considerations may help in the decision-making and define a final value (2).

3- Using the historical evidence of the active comparator and pooling an effect estimate with a 95% CI from the historical RCTs (mostly placebo controlled) (22).

4- The NIM may be determined by the so-called "50%-rule", which advocates that the value of NIM must be inferior (preferentially 50%) to the inferior limit of the 95% CI obtained from historical data that compare control treatment and placebo. Suppose that the difference of 20% between proportions has been obtained from a sample of 200 patients per group and the 95% CI is 11.1-28.2%. Taking the half of the inferior limit

(11.1%), the value suggested to NIM is 5.5%. This method is considered conservative, for it provides a double discount in the margin calculation, thus decreasing the power of the study in demonstrating non-inferiority (2).

5- Traditionally, a NIM has been selected by the size of effects that are considered to be of no clinical relevance or to be outweighed by other benefits of the experimental treatment; this method is called the conventional method (23). Three methods have been suggested: NIM by the conventional method, NIM by the 50% effect retention method and NIM by the 95%-95% method (29).

6- NIM may be determined as a percentage of the control effect estimated for the current study, usually between 10 and 20%. Its definition must, however, take into consideration the therapeutic field and the magnitude of the control group effect; for example, for anti-infectious drugs, more conservative margins are recommended (e.g. 10%) when the expected effect is around 90%, and ampler margins (e.g. 20%) when the anticipated effect is inferior to 80%. The existence of other possible benefits must also be considered; a larger margin is accepted if there are clinical advantages such as an important reduction in adverse effects (2).

7- The point estimate and the fixed-margin methods are methods of analyzing non-inferiority where the margin is defined based either on the effect estimate from the historical evidence or the limit of the CI of the effect estimate that is the closest to the null effect (26).

In the point estimate method, the fraction of the effect estimate that is considered clinically significant is determined based on clinical judgement. This fraction is then called the preserved fraction. The margin is defined based on the effect estimate, which does not capture the variance of the effect estimates of the active comparator from the past trials, and may not reflect the placebo controlled effect of the active comparator that is expected to be present in the non-inferiority trial if a placebo arm were included. This may lead to an estimation of a margin that is either too lenient or too strict (26).

In the fixed-margin method, takes this into account by defining the margin based on the smallest effect size of the active comparator from the past trials (as expressed by the confidence limit that is closest to the null effect). The margin that was defined based on the pooled effect

estimate in the point-estimate method can be used in another method of analysis of NITs which is called the synthesis method. The difference between the point-estimate method and the synthesis method is that in the latter the variance of the effect estimates of the active comparator is incorporated in the analysis of non-inferiority (not in setting the margin) (26).

8- The FDA recommends using the preserved effect method and a fixed margin, which often offers a conservative estimate (30).

First, the margin of the conservative estimate of the entire effect of the standard therapy (M1) is calculated based on historical data (e.g., lower limit of the confidence interval of the effect).

Then, a smaller margin (M2) is selected based on clinical judgment of how much of the standard therapy effect should be preserved.

For example, M2 as 50% of M1 is usually selected for NITs for cardiovascular disease. However, from a scientific point of view, the fixed margin approach suggested by the FDA does not have obvious advantages over the synthesis method, which combines data from the historical study and the current NIT to directly assess the non-inferiority of the experimental therapy without specifying a fixed non-inferiority margin. Furthermore, when the experimental therapy has other important advantages over the standard therapy (e.g., lower risk of severe adverse events), it may be justifiable to accept a wider non-inferiority margin for efficacy (31).

Conclusion

The critical step in determining therapeutic non-inferiority is the selection of NIM. Statistical reasoning and clinical judgment are commonly used to choose this margin. From a statistical perspective, the margin is best determined from a random-effects meta-analysis of historical placebo-controlled studies of the standard treatment (active control). Two key assumptions underlie this determination: 1) the ability to discriminate between effective and ineffective therapies (assay sensitivity or discriminative power) and 2) the applicability of the meta-analysis in the context of the current trial (constancy or representativeness). However, because the non-inferiority trial does not have a placebo control, these assumptions are unverifiable (32).

For this reason, the historical trials being examined should exhibit reliable and consistent superiority of the active control over placebo, and the reference population and the experimental protocol in the current active-control trial should be identical to those used in the historical trials. However, unavoidable inconsistencies in patient characteristics, concomitant medications, intensity of treatment, and temporal improvements in health care can invalidate these key assumptions, thereby rendering previous experience with active control of uncertain relevance to the current study.

Because of this uncertainty, the non-inferiority margin is typically defined in terms of some fraction (f) of the standard treatment effect to be preserved. The choice of f is a matter of clinical judgment governed by the maximum loss in efficacy (the magnitude of inferiority) that one is willing to accept in return for potential non-efficacy advantages of the new therapy. Several factors contribute to this judgment, including the seriousness of the clinical outcome (higher values for death or irreversible morbidity), the magnitude of standard treatment effect (smaller values for large effects), and the overall benefit–cost and benefit–risk assessment.

The choice of margin has a critical impact on sample size (narrower margins resulting in larger numbers) and on statistical uncertainty (inflation of type I error ["false-positive" or erroneous acceptance of an inferior treatment] with wider margins and of type II error ["false-negative" or erroneous rejection of a truly non-inferior treatment] with narrower margins). In summary, several factors contribute to the selection of the margin. Ultimately, the pre-defined selection should be justified on statistical, clinical, and regulatory grounds and should be described explicitly in the published report (32).

Conflict of interest

The author declares that he has no conflict of interest with respect to the research, authorship, and/or publication of this article.

References

1. Fletcher, R. Evaluation of interventions J Clin Epidemiol. 2002;55:1183-90.
2. Pinto, V. Non-inferiority clinical trials: concepts and issues J Vasc Bras 2010;9(3):145-51.
3. Hung, H., Wang, S. Statistical Considerations for Noninferiority Trial Designs Without Placebo Statistics Biopharmaceutical Res. 2013;5(3):239-47.
4. Brown, T.M. Design and interpretation of non-inferiority studies: A clinician's perspective. J Nucl Cardiol. 2017;24(6):1994-7.
5. Wangge, G., Putzeist, M., Knol, M.J., Klungel, O.H., Gispen- De Wied, C.C., de Boer, A. et al.

Regulatory scientific advice on non-inferiority drug trials. PLoS One. 2013;8(9):e74818.

6. Wangge, G., Klungel, O.H., Roes, K.C,, de Boer, A., Hoes, A.W., Knol, M.J. Room for improvement in conducting and reporting non-inferiority randomized controlled trials on drugs: a systematic review. PLoS One. 2010;5(10):e13550.

7. Wiens, B,L. Choosing an equivalence limit for noninferiority or equivalence studies. Control Clin Trials. 2002;23(1):2-14.

8. Rehal, S., Morris, T.P., Fielding, K., Carpenter, J.R., Phillips, P.P. Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. BMJ Open. 2016;6(10):e012594.

9. Piaggio, G., Elbourne, D.R,, Pocock, S.J., Evans, S.J., Altman, D.G,, Group, C. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. JAMA. 2012;308(24):2594-604.

10. Dunn, D.T., Copas, A.J., Brocklehurst, P. Superiority and non-inferiority: two sides of the same coin? Trials. 2018;19(1):499.

11. Schumi, J., Wittes, J. Through the looking glass: understanding non-inferiority. Trials. 2011;12:106.

12. Bouman, A.C., ten Cate-Hoek, A.J., Ramaekers, B.L., Joore, M.A. Sample Size Estimation for Non-Inferiority Trials: Frequentist Approach versus Decision Theory Approach. PLoS One. 2015;10(6):e0130531.

13. Chow, S., Fong, F. On Selection of Margin in Non-Inferiority Trials. J Biom Biostat 2016;7:301.

14. Kaji, A.H., Lewis, R.J. Noninferiority Trials: Is a New Treatment Almost as Effective as Another? JAMA. 2015;313(23):2371-2.

15. Mulla, S.M., Scott, I.A., Jackevicius, C.A, You, J.J., Guyatt, G.H,. How to use a noninferiority trial: users' guides to the medical literature. JAMA. 2012;308(24):2605-11.

16. Pocock, S.J., Clayton, T.C., Stone, G..W. Challenging Issues in Clinical Trial Design: Part 4 of a 4-Part Series on Statistics for Clinical Trials. J Am Coll Cardiol. 2015;66(25):2886-98.

17. Gupta, S.K. Non-inferiority clinical trials: Practical issues and current regulatory perspective. Indian J Pharmacol. 2011;43(4):371-4.

18. Ferreira-Gonzalez, I. Basis for the interpretation of noninferiority studies: considering the ROCKET-AF, RE-LY, and ARISTOTLE studies. Rev Esp Cardiol (Engl Ed). 2014;67(6):432-5.

19. Flandre, P. Statistical methods in recent HIV noninferiority trials: reanalysis of 11 trials. PLoS One. 2011;6(9):e22871.

20. Macaya, F., Ryan, N., Salinas, P., Pocock, S.J. Challenges in the Design and Interpretation of Noninferiority Trials: Insights From Recent Stent Trials. J Am Coll Cardiol. 2017;70(7):894-903.

21. Christensen, E. Methodology of superiority vs. equivalence trials and non-inferiority trials. J Hepatol. 2007;46(5):947-54.

22. Althunian, T.A., de Boer, A., Groenwold, R.H.H., Klungel, O.H. Defining the noninferiority margin and analysing noninferiority: An overview. Br J Clin Pharmacol. 2017;83(8):1636-42.

23. US Food and Drug Administration. Guidance on Statistical Principles for Clinical Trials. International conference on harmonisation; guidance on statistical principles for clinical trials; availability—FDA. Notice Fed Regist 1998;63:449583-98. PMID: 10185190.

24. Rothman, M., Tsou, H. On non-inferiority analysis based on delta-method confidence intervals Biopharm Stat 2003;13:565-83.

25. Head, S.J., Kaul, S., Bogers, A.J., Kappetein, A.P. Non-inferiority study design: lessons to be learned from cardiovascular trials. Eur Heart J. 2012;33(11):1318-24.

26. Althunian, T.A., de Boer, A., Klungel, O.H., Insani, W.N., Groenwold, R.H. Methods of defining the non-inferiority margin in randomized, double-blind controlled trials: a systematic review. Trials. 2017;18(1):107.

27. Hahn, S. Understanding noninferiority trials. Korean J Pediatr. 2012;55(11):403-7.

28. Fleming, T.R., Odem-Davis, K., Rothmann, M.D., Li Shen, Y. Some essential considerations in the design and conduct of non-inferiority trials. Clin Trials. 2011;8(4):432-9.

29. Tanaka, S., Kinjo, Y., Kataoka, Y., Yoshimura, K., Teramukai, S. Statistical issues and recommendations for noninferiority trials in oncology: a systematic review. Clin Cancer Res. 2012;18(7):1837-47.

30. Huitfeldt, B., Hummel, J., European Federation of Statisticians in the Pharmaceutical I. The draft FDA guideline on non-inferiority clinical trials: a critical review from European pharmaceutical industry statisticians. Pharm Stat. 2011;10(5):414-9.

31. Xie, X., Wang, M., Ng, V., Sikich, N. Some issues for the evaluation of noninferiority trials. J Comp Eff Res 2018. 2018;7(9):835-43.

32. Kaul, S., Diamond, G.A. Good enough: a primer on the analysis and interpretation of noninferiority trials. Ann Intern Med. 2006;145(1):62-9.