

# Understanding Superiority, Noninferiority, and Equivalence for Clinical Trials

## Abstract

Randomized control trials are the gold standard for testing the efficacy of new interventions. Historically, superiority trials were methods of choice as reference (standard) interventions were not established for many disease conditions. However currently, reference interventions are available for most of adverse conditions. Despite this, many investigators are using superiority trials in comparison to more suitable noninferiority and equivalence trials. The application of noninferiority and equivalence trials is on the rise, but by and large, these trials are poorly understood, ill-conceived, inappropriately analyzed, and reported and misinterpreted.

**Keywords:** *Clinical trials, equivalence, methodology, noninferiority, superiority*

## Kamal Kishore, Rahul Mahajan<sup>1</sup>

*Departments of Biostatistics and <sup>1</sup>Dermatology, Venereology, and Leprology, Postgraduate Institute of Medical Education and Research, Chandigarh, India*

## Introduction

The efficacy of a drug in randomized control trials (RCTs) is established using null hypothesis significance testing (NHST) approach.<sup>[1,2]</sup> An insightful clinical trial meticulously documents the specific objectives, hypotheses, data analysis, and reporting plans. Traditionally, the investigators were investigating the superiority of the intervention group with a control group in the absence of reference interventions. Therefore, the classical approach was to frame two-sided (there is a significant difference in the outcome variable between two groups) alternative hypothesis. The Consolidated standard of Reporting (CONSORT) for RCTs recommend two-sided *P* value and confidence interval approach to declare the statistical significance.<sup>[3]</sup> After rejecting the null hypothesis, investigators used to compare the average score of the outcome variable between the intervention group (experimental group) and the control group to declare superiority. Recently, many authors are comparing novel interventions (NI) as compared to reference interventions (RI). The competing intervention may not significantly differ in terms of their efficacy. Although, there maybe various other reasons such as safety, low cost, ease of administration,

and lesser side effects of a novel intervention.<sup>[4,5]</sup> Therefore, the traditional approach of superiority trial do not work in these types of situations. The noninferiority and equivalence trials are better methods of investigations for comparing NI against standard treatment. The primary objectives of noninferiority and equivalence trials are to demonstrate that NI is better or not worse than active control (standard intervention or reference intervention) beyond clinically significant margins.<sup>[5,6]</sup> By contrast, noninferiority trials investigate NI (e.g., new drug or treatment claimed to be better than existing standard drug by the company in terms of ease of administration, etc.) is not worse than active control in terms of either or both safety and efficacy beyond a specified margin of clinical significance. However, NI may be better than active control. Equivalence trial investigates novel intervention (e.g., generic drugs claimed by the govt. to have similar efficacy and safety but lower cost as compared to company supplied drug) is not better and worse than active control beyond the prespecified margin of clinical significance. Thus, null and alternative hypothesis and type-I and type-II error will change as per the objectives of the trial. There will also be certain modifications in sample size and statistical analysis plan. The classical approaches of the

**Address for correspondence:**  
Dr. Rahul Mahajan,  
Associate Professor, Department  
of Dermatology, Venereology,  
and Leprology, Postgraduate  
Institute of Medical  
Education and Research,  
Chandigarh - 160 012, India.  
E-mail: [drmahulpgi@yahoo.com](mailto:drmahulpgi@yahoo.com)

### Access this article online

**Website:** [www.idoj.in](http://www.idoj.in)

**DOI:** 10.4103/idoj.IDOJ\_130\_20

### Quick Response Code:



This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

**For reprints contact:** [WKHLRPMedknow\\_reprints@wolterskluwer.com](mailto:WKHLRPMedknow_reprints@wolterskluwer.com)

**How to cite this article:** Kishore K, Mahajan R. Understanding superiority, noninferiority, and equivalence for clinical trials. *Indian Dermatol Online J* 2020;11:890-4.

**Received:** 06-Mar-2020. **Revised:** 09-May-2020.

**Accepted:** 20-Jul-2020. **Published:** 19-Sep-2020.

null hypothesis, sample size, statistical analysis using the  $P$  value are not justifiable for non-inferiority and equivalence trials. Therefore, the framing of appropriate hypotheses is foundations for the application of appropriate superiority, noninferiority, and equivalence trials. Further, it facilitates sample size, statistical analysis, reporting, and conclusions. Initially, we attempt to disentangle the various intricacies of noninferiority and equivalence trial along with familiar superiority trial. Finally, we proposed a brief table of important characteristics to plan and analyses these types of trials.

### Clinically meaningful difference ( $\Delta$ )

The concept of statistical significance and clinical importance is at the heart of clinical research. The clinically relevant difference should always take precedence over statistical significance. A change in score of few points (e.g., 70-65) in the severity of alopecia assessment tool (SALT; range 0-100) among alopecia patients may be statistically significant but may not be clinically relevant.<sup>[7]</sup> Typically, a clinically meaningful difference is determined clinically and statistically after consulting literature, personal experience and discussion with colleagues. It is denoted with symbol  $\Delta$  and is usually challenging to select. A very small  $\Delta$  may lead to rejection of a promising drug, whereas large  $\Delta$  may lead to a selection of potentially inferior drug. Generally, superiority trials justify  $\Delta$  after analysis as compared to equivalence and noninferiority trials. Whereas, a prerequisite knowledge about  $\Delta$  is vital to frame hypotheses for equivalence and noninferiority trials.<sup>[8]</sup> The conclusion about equivalence and noninferiority also depends on  $\Delta$ . It is compelling to specify  $\Delta$  after data inspection, but it may give rise to bias.<sup>[9]</sup> Therefore, it is essential to fix  $\Delta$  with suitable reasons in advance while finalizing research protocol. A conservative rule is to set a clinically significant difference ( $\Delta$ ) of 50% from the value of lower confidence interval in a superiority trial.<sup>[6]</sup> The FDA recommended the  $\Delta$  value of 20, 15, 10, and 5% for the anti-infective drugs with the efficacy of 50-80, 80-90, 90-95, and >95%, respectively, measured as the binary outcome.<sup>[10]</sup>

### Hypothesis and error

An initial step in almost all the studies is to formulate the hypothesis. A superiority trial investigates whether NI is better than active control (AC) or placebo by a specified margin ( $\Delta$ ). The null and alternative hypothesis for superiority trials are  $H_0: \mu_{NI} - \mu_{AC} \leq \Delta$  and  $H_1: \mu_{NI} - \mu_{AC} > \Delta$ , respectively. However, in practice, a superiority trial is a two-step NHST process. Therefore, as per NHST, null and alternative hypothesis are formulated as  $H_0: \mu_{NI} - \mu_{AC} = 0$  and  $H_1: \mu_{NI} - \mu_{AC} \neq 0$ . After rejecting the null hypothesis, the adequacy of  $\Delta$  is determined while reporting results. The authors can read an excellent article written by Farrugia *et al.* on formulating the research question, hypotheses and objectives.<sup>[11]</sup> The equivalence trial investigates whether NI

is equally good as compared to AC. The active control is the standard intervention for noninferiority and equivalence trials as compared to placebo or non-standard intervention for superiority trial. The equally good margin ranges from  $-\Delta$  to  $+\Delta$  and known as tolerance range for equivalence studies. The noninferiority trials examine whether NI is not noninferior to AC. The NI will be noninferior to AC (standard intervention) if the noninferiority margin is to right of  $-\Delta$ . The type-I error is critical for conclusions. Type-I error is falsely rejecting a null hypothesis (i.e., falsely declaring statistical significance). The interpretation is inherently confusing to investigators in general for equivalence and noninferiority trials. The objectives of null and alternative hypotheses for superiority (rejection of null) and equivalence trials (rejection of difference) are opposite to each other. Whereas alternative hypothesis for noninferiority trial is one-tailed as compared to two-tailed for superiority trials. Researchers can refer Table 1 for a lucid explanation of hypotheses and errors for the different type of trials.

### Sample size

The optimum sample size is one of the crucial requirements to conduct systematic and objective research. It is unethical and unscientific to conduct underpowered studies. Typically, there are four fundamental requirements, such as variance, confidence interval (CI), power, and effect size to calculate the sample size. However, it is essential to prespecify with an appropriate explanation for noninferiority and equivalence trials. Table 2 gives the formula for calculating sample size for continuous and binary outcome variables in different types of clinical trials. The researcher interested in detailed discussion regarding sample size for many other types of designs can consult "Sample Size Calculations in Clinical Research" by Chow *et al.*<sup>[10]</sup> The total sample size ( $n$ ) is sum ( $n_1+n_2$ ) of sample sizes in each arm. The value of  $k = 0.5, 1$  and  $2$  represent 1:2, 1:1, and 2:1 participants in treatment and control group, respectively. Typically, the value of  $k$  is taken as zero for superiority trials as the null hypothesis is usually framed as a hypothesis of no difference. In general, sample size to conduct noninferiority and equivalence studies are higher as compared to superiority trials.<sup>[9]</sup> Let us assume the effectiveness of achieving an intended endpoint (say PASI75) with a placebo and a drug under investigation is 25% and 60%, respectively. Normally,  $d = 0$  for superiority trials. The  $z$  and values at 95% CI and 80% power are 1.96 and 0.84, respectively. The values of  $P_1 = 0.25$  and  $P_2 = 0.60$  and  $d = 0.25-0.60 = -0.35$ . Assuming, investigator need equal number of cases to controls ( $K = 1$ ). The application of sample size formula for superiority trial from Table 2 give a sample of approximately 29 individuals to be recruited in each group. Similarly, assume the effectiveness of achieving PASI75 for noninferiority trial with a standard drug and novel interventions are 55% and 60%, respectively. The  $z$  and values at 95% CI and 80% power are 1.64 and 0.84, respectively. The values

of  $P_1 = 0.55$  and  $P_2 = 0.60$  and  $d = 0.55 - 0.60 = -0.05$ . A noninferiority margin of 10% ( $= -0.10$ ) is considered meaningful for the outcome. Assuming, investigator need equal number of cases to controls ( $K = 1$ ). The application of sample size formula for noninferiority trial from Table 2 give a sample of approximately 134 individuals to be recruited in each group. Similarly, assuming researcher is interested in equivalence trial for the measures given for noninferiority trials. All the values except values at 95% CI and 80% power are 1.96 and 1.28 will remain the same. The application of sample size formula for equivalence trial from Table 2 give a sample of approximately 2047 individuals to be recruited in each group.

### Statistical analysis

The CONSORT guidelines recommend using the  $P$  value and CI approach for declaring the superiority of one intervention over others.<sup>[3]</sup> However, many studies still report results using the  $P$  value approach. The superiority using CI can be declared if the CI does not include 0 (1 for odds and risk ratio). When it comes to making inference about non-inferiority and equivalence trials, CI approach is recommended in comparison to the  $P$  value approach. The intention to treat (ITT) and

per-protocol analysis have advantages and disadvantages. ITT is the preferred analytic approach for reporting results from superiority trials in comparison to both ITT and per-protocol analysis for equivalence and noninferiority trials.<sup>[6,12]</sup> Usually, both lead to the same conclusion. However, in case of disagreement, a judicious and careful approach after consulting literature needs to be adopted. The clinically significant difference ( $\Delta$ ) plays an essential role in declaring non-inferiority and equivalence. The CI to declare equivalence lies between  $-\Delta$  and  $+\Delta$ . For the noninferiority trial, novel intervention needs to be similar or better than reference intervention. Therefore, the CI to declare non-inferiority lies on the right side of  $-\Delta$ . Figures 1 and 2 display the confidence interval approach to declare significance for positive (higher the better) and negative (lower the better) outcome variable, respectively, for various trials.

### Contextualizing

#### Scenario

Understanding the choice of trial design based on the evolution of first-line systemic therapies in severe psoriasis from methotrexate to biologics and biosimilars.

**Table 1: The hypotheses and errors for different type of clinical trials**

Type of trial	Null hypothesis	Alternative hypothesis	Type-I error	Type-II error
Superiority	There is no significant difference between novel intervention and Placebo	There is a significant difference between the novel intervention and placebo	Erroneously concluding superiority of novel intervention when there is no superiority	Not concluding a superiority when there is a superiority
Noninferiority	The novel intervention is not noninferior to reference intervention	The new intervention is noninferior to reference intervention	Erroneously concluding noninferiority of novel intervention when there is no noninferiority	Not concluding noninferiority when there is noninferiority
Equivalence	The novel intervention is not equivalent to reference intervention	The novel intervention is equivalent to reference intervention	Erroneously concluding equivalence of interventions when there is no equivalence	Not concluding equivalence when there is equivalence

**Table 2: Sample size formulae for different type of trials for continuous and binary outcome variable**

	Continuous outcome variable	Binary outcome variable
Superiority trial	$n_1 = \left[ \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \sigma^2 (1 + \frac{1}{k})}{(d - \Delta)^2} \right], n_2 = 1$	$n_1 = \left[ \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \left[ \frac{P_1(1 - P_1)}{K} + P_2(1 - P_2) \right]}{(d - \Delta)^2} \right], n_2 = kn_1$
Noninferiority trial	$n_1 = \left[ \frac{(z_{1-\alpha} + z_{1-\beta})^2 \sigma^2 (1 + \frac{1}{k})}{(d - \Delta)^2} \right], n_2 = kn_1$	$n_1 = \left[ \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 \left[ \frac{P_1(1 - P_1)}{K} + P_2(1 - P_2) \right]}{(d - \Delta)^2} \right], n_2 = kn_1$
Equivalence Trial	$n_1 = \left[ \frac{(z_{1-\alpha} + z_{1-\beta/2})^2 \sigma^2 (1 + \frac{1}{k})}{(\Delta -  d )^2} \right], n_2 = kn_1$	$n_1 = \left[ \frac{(z_{1-\alpha/2} + z_{1-\beta/2})^2 \left[ \frac{P_1(1 - P_1)}{K} + P_2(1 - P_2) \right]}{(\Delta -  d )^2} \right], n_2 = kn_1$

$z_{1-\alpha/2} = 1.64, 1.96, \text{ and } 2.58$  for 90%, 95% and 99% for 2-sided confidence interval respectively,  $z_\beta = 0.84$  and  $1.28$  for 80% and 90% power respectively,  $\sigma^2 =$ Expected variability (obtained from previous studies).  $P_1 =$  Proportion of participants with outcome in the first group,  $P_2 =$ Proportion of participants with outcome in the second group.  $k =$ Allocation ratio between treatment and control group,  $d =$ Effect Size,  $\Delta =$ Clinically meaningful difference

The relevance of understanding the trial design for superiority, inferiority, or equivalence can be understood in context to the gradual emergence of biologic response modifiers in the management of severe psoriasis. At the turn of the century, there were limited data on the clinical efficacy and adverse effect profile as targeted therapies were emerging for treating severe psoriasis. The early randomized controlled trials were placebo-controlled.<sup>[13]</sup> It leads to establishing of methotrexate as a standard drug. Subsequently, a three-arm noninferiority trial popularly known as CHAMPION trial compared adalimumab (NI) with the gold standard drug methotrexate,<sup>[14]</sup> and a placebo. Despite being designed as a non-inferiority trial, the CHAMPION trial declared the superiority of adalimumab as compared to methotrexate. The Committee for Proprietary Medicinal Products (CPMP) have discussed conditions under which results of noninferiority trial can be reported as a superiority trial.<sup>[9]</sup>

As mentioned above, placebo-controlled superiority trials were the first steps for new drugs to show its potential

benefits. The recommendations are for noninferiority designs to compare the new drug (adalimumab in this case) relative to an established one (methotrexate here). The CHAMPION trial was designed as a double-blind, double-dummy placebo-controlled study with an adequate sample size of 90% power to detect noninferiority of adalimumab relative to methotrexate (based on the primary outcome of achieving PASI75- assuming effectiveness of adalimumab to be 62% and of methotrexate to be 60%, and placebo 4%).<sup>[14]</sup> Despite few limitations or criticisms of the trial, CHAMPION trial proved to be a vital trial in turning the tide toward the more widespread use of biologics in comparison to the traditional systemic agents in managing moderate to severe psoriasis. Recently, many biosimilars came into the market after the expiry of patent for adalimumab (Humira) in both the United States and Europe. Hercogova *et al.* compared the efficacy, safety, and immunogenicity of the biosimilar MSB11022 (a new agent here) with reference adalimumab (now an established agent).<sup>[15]</sup> Before the clinical trial, the biosimilar was

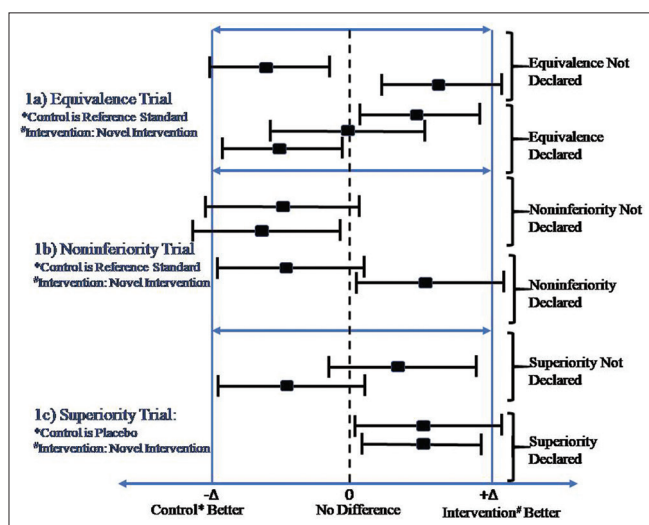


Figure 1: Interpretation of results from confidence intervals with different type of trials for positive outcome

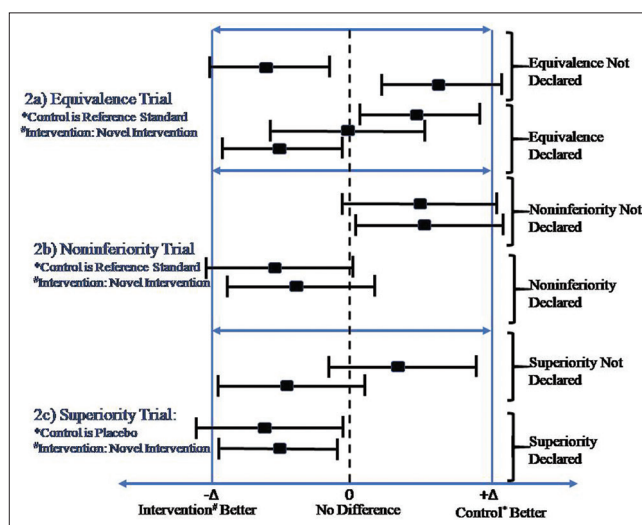


Figure 2: Interpretation of results from confidence intervals with different type of trials for negative outcome

Table 3: Summary of important characteristics of difference between different type of trials

	Type of trial		
Characteristics	Superiority trial	Noninferiority trial	Equivalence trial
Condition for application	Comparing novel intervention with non-standardized intervention	Comparing novel intervention with standard intervention	Comparing novel intervention (e.g., generic drug) with standard intervention
Control	Placebo or non-standard Intervention	Standard intervention	Standard intervention
Intervention	Novel intervention	Novel intervention	Novel intervention
Hypothesis	Two-tailed	One-tailed	Two-tailed
Key to hypothesis formation	Effect size (d)	Effect size (d) and clinically meaningful difference ( $\Delta$ )	Effect size (d) and clinically meaningful difference ( $\Delta$ )
Statistical Significance range	$\mu_1 - \mu_0 \neq 0$	$\mu_1 - \mu_0 > -\Delta$	$-\Delta \leq \mu_1 - \mu_0 \leq \Delta$
Analysis recommendation	ITT	ITT and per protocol analysis	ITT and per protocol analysis
Reporting	P-value and CI	CI	CI

CI - Confidence intervals; ITT - Intention to treat

shown to have structural, functional, and pharmacokinetic equivalence to adalimumab. The researchers can consult Table 3 as a ready reference to decide about the selection of one trial over another.

## Conclusion

By default, superiority trials are preferred by researchers. However, the usage and reporting of noninferiority and equivalence trials are increasing. These trials are more complex and challenging to understand and interpret as compared to superiority trials. The researchers need to better equip themselves with the intricacies and subtle differences between various trials to keep themselves abreast with the latest developments. Careful attention will help researchers to make an informed decision about the claimed safety and efficacy of interventions.

## Financial support and sponsorship

Nil.

## Conflicts of interest

There are no conflicts of interest.

## References

- Sackett DL. Superiority trials, non-inferiority trials, and prisoners of the 2-sided null hypothesis. *Evid Based Med* 2004;9:38 LP-39.
- Wang B, Wang H, Tu XM, Feng C. Comparisons of superiority, non-inferiority, and equivalence trials. *Shanghai Arch Psychiatry* 2017;29:385-8.
- Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Jpn Pharmacol Ther* 46.
- 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;18:1837-47.
- Greene CJ, Morland LA, Durkalski VL, Frueh BC. Noninferiority and equivalence designs: Issues and implications for mental health research. *J Trauma Stress* 2008;21:433-9.
- Head SJ, Kaul S, Bogers AJC, Kappetein AP. Non-inferiority study design: Lessons to be learned from cardiovascular trials. *Eur Heart J* 2012;33:1318-24.
- Olsen EA, Hordinsky MK, Price VH, Roberts JL, Shapiro J, Canfield D, *et al.* Alopecia areata investigational assessment guidelines—Part II. *J Am Acad Dermatol* 2004;51:440-7.
- Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG. CONSORT Group, *et al.* Reporting of noninferiority and equivalence randomized trials: Extension of the CONSORT 2010 statement. *JAMA* 2012;308:2594-604.
- Committee for Proprietary Medicinal Products (CPMP). Points to consider on switching between superiority and non-inferiority. *Br J Clin Pharmacol* 2001;52:223-8.
- Chow S-C, Shao J, Wang H, Lohknygina Y. *Sample Size Calculations in Clinical Research*. New York: Chapman and Hall/CRC; 2017.
- Farrugia P, Petrisor BA, Farrokhyar F, Bhandari M, *et al.* Practical tips for surgical research: Research questions, hypotheses and objectives. *Can J Surg* 2010;53:278-81.
- Pater C. Equivalence and noninferiority trials - Are they viable alternatives for registration of new drugs? (III). *Curr Control Trials Cardiovasc Med* 2004;5:1-7.
- Gordon KB, Langley RG, Leonardi C, Toth D, Menter MA, Kang S, *et al.* Clinical response to adalimumab treatment in patients with moderate to severe psoriasis: Double-blind, randomized controlled trial and open-label extension study. *J Am Acad Dermatol* 2006;55:598-606.
- Saurat JH, Stingl G, Dubertret L, Papp K, Langley RG, Ortonne JP, *et al.* Efficacy and safety results from the randomized controlled comparative study of adalimumab vs. methotrexate vs. placebo in patients with psoriasis (CHAMPION). *Br J Dermatol* 2008;158:558-66.
- Hercogova J, Papp KA, Chyrok V, Ullmann M, Vlachos P, Edwards CJ. AURIEL-PsO: A randomized, double-blind phase III equivalence trial to demonstrate the clinical similarity of the proposed biosimilar MSB11022 to reference adalimumab in patients with moderate-to-severe chronic plaque-type psoriasis. *Br J Dermatol* 2020;182:316-32.