Statistical Considerations for Randomized Controlled Trials [RCTs] - Understanding Superiority, Equivalence and Non-inferiority Designs

NJ Gogtay, S Deshpande, UM Thatte

Introduction

Randomized controlled studies (RCTs) are considered the gold standard for answering research questions related to the effect of one treatment (be it a drug, device, vaccine, diagnostic test, life-style intervention like yoga) relative to another. When these interventions are investigated, there are three potential ways in which we can compare them. Let us understand this with an example. A new anti-hypertensive drug is to be studied before it is marketed. We would either compare it to a placebo or an existing anti-hypertensive drug. This comparison could hypothesize one of the three scenarios described below:

1. Better than the existing drug or placebo?
2. As effective as the existing one?
3. No worse than the existing one?

The first hypothesis is the one that we are most familiar with and the “effect” that we are interested in is one of “superiority” [is the new drug better than the existing one?]. Thus, in clinical practice, the new drug can replace the old anti-hypertensive as the standard of care, if indeed it is found to be better than the old one. This is well illustrated by studies that were designed to show artemisinin derivatives to be superior to quinine for the management of severe P. falciparum malaria. These studies of superiority changed clinical practice with artemisinin replacing quinine in the treatment of this infection.

However, recognizing the fact that it is becoming increasingly difficult to develop newer therapies with better [superior] efficacy than the current existing treatments [which have become standards of care], the science of drug development has witnessed the emergence of two equally important designs. These are the “equivalence” and “non-inferiority” designs, which aim to answer the next two questions mentioned above (is the new drug as effective as the existing one and is it no worse than the existing one, respectively).

While the results of studies with superiority design eventually change clinical practice by replacing the standard of care with the new drug as described, designs of equivalence and non-inferiority aim to make possible the “exchange or swap” of the new intervention (most commonly for reasons such as lower cost, reduced adverse effects, or greater ease of administration, among others) for the existing treatment, thus increasing options for clinical management.

All three designs have unique statistical requirements and this paper will discuss these along with key design features.

1. The Superiority trial: The superiority study design is essentially needed to achieve the objective of establishing a definitive “difference” between the two interventions with the new being better than the older treatment. This design is often used during early stages of new drug development, where it is critical to show that the new drug is clearly superior to placebo [or the vehicle or excipient in which it is formulated] as not showing this will mean that the new drug is simply not effective and not worth further development.

The null hypothesis in a superiority design would be that the two treatments are “not different” and the alternative hypothesis that the new intervention is better (superior) than the existing treatment.

The first step when planning a superiority trial is to define the magnitude or the quantum of difference that we will accept between the new treatment and the existing treatment as...
If we were to continue with control/old treatment better, we may state a “single value” for this margin, at the end of the study we will actually get a range of values where the true difference between the two treatments would eventually lie. This is the 95% Confidence Interval [95% CI] and for a superiority trial, the lower bound of this 95% CI should lie ABOVE the superiority margin that has been set. If we were to continue with our example of the two anti-hypertensive drugs [see above], and if we wish to show that anti-hypertensive B is “superior” to anti-hypertensive A; we need to first “fix” the superiority margin, which we will do based on either pilot studies, published literature or animal studies. Although it remains a “guestimate” it should be based on as much evidence as is available at that point of time for reasonable results. In our example, say we anticipate the mean difference [superiority margin or effect size] between the two drugs to be 10mm Hg. At the end of the study if we get a 95% CI of, then we have proven superiority. If we got a 95% CI [8,55] our drug would have failed to show superiority as the lower bound of the 95% CI is falling below the set superiority margin.

Another point to note is that in superiority designs, the 95% CI does not include zero! This is because if zero were to fall in this range, it would mean there was “no difference” between the treatments A and B!

Thus, a study with the objective of establishing superiority must satisfy the following two statistical requirements (Figure 1).

**a.** At 5% significance, the mean difference between the two treatments is significantly different [reduction in blood pressure in this case] and

**b.** The 95% CI of the mean difference EXCLUDES zero

**2.** **Equivalence trials:** An equivalence trial is conducted when we aim to show that the “newer” or often called the “test” intervention is “as effective” as the reference or comparator intervention. The desired result here is that while the two treatments being compared may show some differences, these are “not unacceptably different” or the treatments are “clinically indifferent” and hence one can easily be substituted for the other. The term equivalent is not used here in the strictest sense of the word, but rather in the context of efficacies that are so close that neither intervention can be considered superior or inferior to the other.

The first step when planning an equivalence trial is to define the range of the “acceptable” difference between the new treatment and the existing treatment. This is called the “equivalence margin” and it should be defined or set prior to starting the study. The next step would be to define the 95% CI of the mean difference in which the final result should lie.

Let us take the example of a recent study conducted at the district level in Uganda compared the effectiveness of misoprostol given by a midwife (comparator) to that given by a physician (standard of care) to treat women who had an incomplete abortion. The objective of the study was to show that misoprostol given by midwives was “equivalent” or “not unacceptably different” from that given by physicians. The predefined “equivalence range” was -4% to 4% expressed as proportions of women who needed surgical intervention.

At the end of the study, the mean difference found between the two groups was -0.8% with 95% CI [-2.9, 1.4]. This mean difference and both bounds of the 95% CI fell within the pre-specified equivalence margins of -4% and 4% indicating that treatment of an incomplete abortion by midwives with misoprostol was “not unacceptably different” than or “equivalent” to the treatment by physicians.

Thus a study with the objective of establishing equivalence needs to satisfy the following two requirements (Figure 2).

**a.** At 5% significance, the mean difference between the two treatments is not significantly different and the 95% CI of the

![Fig. 1: Superiority margin and the relationship between the p value and 95% CI for a superiority trial that compares an new treatment to an older one](image)
mean difference lies within the pre-specified equivalence margins
and
b. The 95% CI of the mean difference INCLUDES zero: Bioequivalence studies are a special case of equivalence studies where pharmacokinetic profiles of the generic drug is compared to the innovator drug to show “equivalence” with respect to certain prominent pharmacokinetic properties of both drugs such as maximum concentration achieved [Cmax], time taken to reach the maximum concentration [Tmax] and Area under the curve [AUC]. Since bioequivalence studies compare the innovator drug with a generic one, this permits substitution of the innovator drug with the generic thereby leading to cost savings for the payer.

3. The Non-inferiority trial: Non-inferiority trials aim to determine if a test or new therapy is no worse than a standard therapy for a particular outcome of interest. Here, in contrast to equivalence, the aim is to show that a new product is not unacceptably worse than an older one. However, similar to an equivalence trial, we accept some (pre-defined) inferiority of the test intervention for specific reasons such as- the new treatment may be cheaper, given fewer times a day, have a more convenient dosing schedule, have fewer side-effects or even simply lead to a better quality of life for the patient. An example of this would be the development of combination vaccines, where several vaccines are given together in a single injection. When such a strategy is devised, it needs to be shown that the single shot of the combination vaccine (in whatever schedules recommended) is non-inferior to giving all vaccines singly in terms of antibody titers.

However, here, the null hypothesis peculiarly seems backwards! The null hypothesis thus states that the new treatment is “worse” than the old by a certain “margin” while the alternate hypothesis states that the difference in effect between the two treatments is “less” than this margin.

The first step when planning a non-inferiority trial is to define the magnitude or quantum of inferiority that we will accept between the new treatment and the existing treatment as both statistically significant and clinically meaningful. This magnitude is also called the “non-inferiority margin” and like in the superiority and equivalence designs, it should be defined à priori before starting the study. Like with a superiority trial, we may state a “single value” for this margin, at the end of the study we will actually get a range of values where the true difference between the two treatments will lie and the lower bound of this 95% CI should lie ABOVE the non-inferiority margin that has been set.

Take for example, a study conducted in Thailand that evaluated whether CD4 count based monitoring and switching strategy that is used in low and middle income countries was non-inferior or no worse than viral load based monitoring practiced in developed countries. The primary outcome measure was the proportion of patients with clinical failure in the two groups and the non-inferiority margin was set at -7.4%. At the end of the study, the mean difference for the primary outcome measure was -0.6% with 95% CI of -4.5% to 3.4% proving the non-inferiority of CD4 based monitoring vis a vis viral load monitoring for antiretroviral treatment.

Thus, a study with the objective of establishing non-inferiority satisfies the following two requirements (Figure 3).

a. At 5% significance, the mean difference between the two...
4. Sample size considerations:
   a. The 95% CI of the mean difference includes zero.
   b. The 95% CI of the mean difference includes zero and the non-inferiority margin.

   Fig. 3: 95% CI and the non-inferiority margin. Adapted from Walker E & Nowacki AS.

For example, the COBALT study that compared continuous infusion of alteplase with double bolus alteplase administration for acute myocardial infarction studied 7169 patients using a non-inferiority margin of 0.4%. If an equivalence design was used, the number of patients required to show the treatments to be similar would be 50,000 and if a superiority design was to be used to show a reduction in mortality from 10 to 7.5% [a difference of 2.5%], only 4000 patients would be needed.

5. Suggestions to keep in mind when using non-inferiority and equivalence study designs:
   
   i. It is important to set the margins for non-inferiority and equivalence studies based on a combination of clinical judgment and statistical reasoning. If a very wide margin is set non-inferiority or equivalence may be proven at the end of the study, but the patients will eventually have to bear the burden of an inferior comparator. Hence, the choice of the margin should be ethical and scientifically sound and the conduct of these studies of high methodological quality and rigor.
   
   ii. Similarly, the choice of the comparator is equally important as an “unfair” or poorly effective comparator is used, then showing non-inferiority or equivalence to this would result in a poor product being accepted in clinical practice.
   
   iii. Although the first papers of the non-inferiority and equivalence designs can be traced back to the late seventies, there has been considerable evolution of these designs since then including guidelines [CONSORT extensions] on how these studies are to be reported in literature.

References