Non-inferiority and equivalence trials: Need for a standardized process

In the initial period of 1900, or even a little before, the efficacy of any new treatment used to be assessed on the basis of information gathered through the medical report of one patient or several patients (case series). In very few studies, the efficacy of treatment used to be very obvious, even in a small sample of patients. In the rest (relatively large number) of the studies, it was not that easy to draw a meaningful conclusion with regard to the efficacy of treatment, as the number of patients exposed to the treatment was relatively less. Thereafter, began a period of studies designed to evaluate the efficacy of treatment using a large number of patients. Bradford Hill noticed that all these studies were uncontrolled, and came out with the idea of conducting randomized clinical trials (RCT) in the field of medicine.[1] Thus, for a long period of time, RCT continued to be the best method for comparing the effects of two treatments.[2,3] The objective of the RCT was to show the superiority of one treatment over the other — in most cases, the other treatment used to be placebo.

Showing superiority of a new treatment or test treatment (TT) over placebo (P) or conducting ‘placebo-controlled’ trials has been the gold standard for many years in drug development.[4] However, with the increased availability of established treatments with proven efficacy, the placebo-controlled trials started facing ethical issues. This gave rise to an era of ‘active-controlled’ trials and the focus shifted from showing TT being superior to P to showing TT ‘as good as’ the standard treatment (ST). Statistically, it was not possible to show exact equality in the efficacy of the two treatments, hence, it was demonstrated that the effect of TT was ‘as good as’ ST. Subsequently, the phrase ‘as good as’ got replaced by the word ‘equivalence’ and the category of ‘equivalence trial’ came into existence.

Statistical tests used in superiority are called superiority trial tests. If the superiority trial test is significant, one concludes that the efficacy of TT is different from that of ST. Furthermore, if the result is in favor of TT, the conclusion is that TT is statistically and significantly superior to ST. However, a nonsignificant superiority test is misinterpreted as an indirect evidence of ‘no difference’ between the two treatments or ‘equivalence of the two treatments’. Unfortunately as per the statistical principles of hypothesis testing, if the statistical test implies that the difference is not the statistically significant, then it cannot be concluded that the two treatments are ‘equivalent.’ Hence, use of apt statistical principles is very important for setting up a null hypothesis for ‘equivalence’ trials or ‘non-inferiority (NI)’ trials.

It is well known that the trials using non-inferiority designs are being conducted since 1982,[5] and many products have been approved by regulatory agencies based on the results of studies conducted using non-inferiority design.[6] However, the ‘Draft Guidance for Industry: Non-Inferiority Clinical Trials’ was distributed for comment purposes only, in March 2010, by the US Food and Drug Administration (FDA).[7]

Regulations on adequate and well-controlled studies (21CFR 314.126) describe the following four types of concurrently controlled trials for providing evidence of effectiveness:

- Placebo-controlled
- No treatment
- Dose-response
- Active treatment (active–controlled)

The first three types mentioned above are superiority trials, and attempt is to show that TT is superior to the control (placebo, no treatment or a lower dose of TT). The fourth active-controlled type can also be categorized as a superiority trial, if the objective of active-controlled trial is to show that TT is more effective than the active–control / standard treatment (ST). Generally, the difference between two active treatments (TT and ST) is less than that between active treatments ST and P or TT and P, hence, achieving statistical significance with the same number of study subjects may fail to show superiority. Hence, the intention...
of such a trial is not to show superiority, but to show that TT is not worse than the comparator, which typically is the standard treatment (ST) or an active–control.

Three important reasons for conducting active control trials are:

- Assay sensitivity: Assay sensitivity is a property of a clinical trial defined as the ability to distinguish an effective treatment from a less effective or ineffective treatment.[8] It should be appropriately used in the context of NI trials
- Ethics: It is unethical to use placebo as a control when established treatments with proven efficacy are available in the market
- Comparative evaluation: This is required to examine how TT compares with ST (which can be an active treatment) available in the market. ST must be known to be effective in the population under study.

The purpose of active control can be to show that a TT is either
- Superior to the active control or
- Equivalent to the active control or
- Non-inferior to the active control

The objective of equivalence trials[9] is to determine whether or not the TT is therapeutically similar to the existing ST (active control), whereas, the objective of NI trials is to determine whether or not the TT is no worse than the existing ST. As mentioned earlier, it is impossible to prove the exact equality, hence, a margin (Δ or M) of non-inferiority for a primary endpoint is defined in advance. NI or the equivalence margin Δ(or M) is the degree of acceptable inferiority between the TT and the ST, where a trial needs to predefine at the design stage. The margin (Δ or M) chosen should be smaller than the ‘effect size’. The effect size is defined as the expected difference between ST and P. It is a reliable and realistic estimate of the ‘difference’ between the effects of ST and P.

As of today, many non-inferiority trials have already been conducted and are being conducted, yet there are many areas that need the attention of the stakeholders of clinical trials as well as regulatory bodies. The basic idea is to sit together and bring about clarity in terms of various requirements, in order to draw a meaningful conclusion, scientifically acceptable to all. Some of the requirements needing immediate attention are listed herewith:

**DESIGN: THREE ARM TRIALS OR TWO ARM TRIALS**

**Three arm trials**

i. In three arm trials, the three arms are: TT, ST, and Placebo (P)

ii. This design will be useful to provide information on the superiority of TT and ST over placebo, which facilitates the defining of the NI margin (Δ or M)

iii. In three arm trials, TT must be shown to be statistically and significantly superior to P. This means that the lower bound of the 95% confidence interval (CI), for the difference between TT and P (TT-P), must be above zero. (then clinical judgment is used to check if the observed value for the difference is clinically relevant)

iv. In such a three arm trial, if both TT and ST fail to show statistically significant superiority over P, it means the trial lacks assay sensitivity

**Two arm trials**

i. In two arm trials, the arms are: TT and ST. There is no placebo or P arm

ii. Unlike the three arm trial, in a two arm trial, we will not have any idea of how TT compares with P

iii. Due to the absence of the placebo (P) group, results of the previous studies comparing ST with P will be required to establish that ST has efficacy

iv. In such a situation, the challenge is to define or set up the NI margin (Δ or M) in advance

v. Thus, to set up the inferiority margin (Δ or M), it is essential to take the support of previous studies from literature, which have compared the ST with P in the intended patient population. The choice or selection of the studies for this purpose has to be done carefully, in order to avoid the following potential disturbances

1. Selection bias: This may crop up if the criteria for selection of suitable studies is not predefined and documented
2. Lack of constancy of trial design over a period of time: Like change in entry criteria, methods
3. Non-uniformity in clinical practice over a period of time
4. Non-uniformity of effects over a period of time
5. Publication bias — studies with positive outcomes are more likely to get published than others with negative results, leaving behind only positive trials. This bias has a risk of missing some realistic and valuable findings in the unpublished negative trials

**CONDUCT OF PAST TRIALS USED TO DEFINE (Δ OR M)**

All earlier studies, mentioned above, comparing ST with P, must match closely in terms of features discussed earlier, like constancy over a period of time, uniformity in all respects, including adherence to protocol, dropout, and incorrectly recruiting patients not likely to respond. The NI margin (Δ) must be predefined in the protocol, to maximize the validity of the procedure. Selection of Δ is a clinical issue and not a statistical one, it has to be clinically relevant and can be decided in consultation with other concerned scientists and if required, even experts from the regulatory agencies. A thumb rule followed is to have Δ equal to a certain percentage of ST’s effect over P (ST-P). For example, it can be 80% of the effect (ST-P). A fixed value can also be chosen with appropriate scientific justification.

**ANALYSIS**

Analysis of the population plays an important role in the statistical analysis of NI trials.

In superiority trials, the intention to treat (ITT) analysis (that is analyzing all randomized subjects, regardless of whether they completed the allocated treatment) is recommended. Some research studies reveal that ITT analysis results in a smaller value (but not always, as found by Brittain and Lin) of treatment difference, if all subject had adhered to the treatment. In case it is a smaller value then there is a risk of falsely claiming NI. Hence, while planning NI studies, appropriate consideration has to be given to the type of patient population to be included in the statistical analysis.

**INTERPRETATION**

There exists a modified hypothesis testing framework for effective analysis of data. Yet a more informative CI approach is preferred in all activities of NI and equivalence trials.

Interpretation of the equivalence and NI trial results depends on the position of the CI for the treatment effects in relation to both null effect (zero difference) and inferiority margin (Δ or M). Thus, the observed treatment difference alone is not sufficiently informative.

In order to claim equivalence it is essential that the CI falls wholly between (Δ or M) and + (Δ or M), as shown in Figure 1.

Similarly, the CI approach for claiming NI will be as shown in Figure 2.

Thus, interpretation of the NI trial needs a thorough consideration of the impact of all the aspects discussed earlier, assumptions underlying the designs, tests used, and placement of the 95% CI for the treatment effect in relation to both — (i) the margin of non-inferiority and (ii) Null or no effect as indicated.

---

**Figure 1:** Confidence interval approach to claim equivalence. ST: Standard treatment, TT: Test treatment

**Figure 2:** Confidence interval approach to non-inferiority. ST: Standard treatment, TT: Test treatment
It is important to note that Figures 1 and 2 presented above are generated only after completing a set of important activities in the clinical trial process. These activities include planning, design, conduct, analysis, and interpretation of results. Any non-scientific thinking in each of these activities or any weak link in the chain can result in drawing an incorrect conclusion, leading to the development of inferior drugs and causing harm to the society. Hence, it is the right time to short list the scientific principles and focus on deriving a standardized procedure for conducting NI trials.

Suresh Keshav Bowalekar
Managing Director, PharmaNet Clinical Services Pvt. Ltd., Mumbai, India

Address for correspondence:
Dr. Suresh Keshav Bowalekar,
Managing Director, PharmaNet Clinical Service Pvt. Ltd., Marwah Center, 7th Floor, Krishanlal Marwah Marg, Andheri (East) – 400 072, Mumbai, India.
E-mail: sbowalekar@pharmanet.com

REFERENCES

1. Hill AB. Statistical Methods in Clinical and Preventive Medicine. New York: Oxford University Press; 1962.
2. Armitage P, Berry G. Statistical Methods in medical Research. 3rd ed. Oxford: Blackwell; 1994.
3. Pocock SJ. Clinical Trials: A practical approach. Chichester: Willy; 1983.
4. Hwang IK, Morikawa T. Design issues in non-inferiority / equivalence trials. Drug Inf J 1999;33:1205-18.
5. Temple, RJ. FDA experience and perspective on non-inferiority trials: Presentation at FDA workshop on CAP, 2008. Available from: http://www.fda.gov/downloads/downloads/Drugs/DraftGuidance/InformationbyDrugClass/UCM187447.pdf [Last accessed on 2011 Sep 8].
6. Deng, CQ. Non-inferiority clinical trials – now comes FDA’s draft guidance. In:On biostatistics and clinical trials. Available from: http://Onbiostatistics.Blogspot.Com/2010/2010/03/Non-Inferiority-Clinical-Trials-Now.Html [Last accessed on 2011 Aug 24].
7. Guidance for industry: Non-inferiority clinical trials – Draft guidance, 2010, March. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf [Last accessed on 2011 Aug 24].
8. ICH Harmonised tripartite guideline: Choice of control group and related issues in clinical trials - E10: Current Step 4 version dated 20 July 2000. Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E10/Step4/E10_Guideline.pdf [Last accessed on 2011 Aug 24].
9. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials – An extension of the CONSORT statement Available from http://jama.ama-assn.org/content/295/10/1152.full [Last accessed on 2011 Aug 24].
10. Hopewell S, Clarke M, Moher D, Wager E, Philippa M, Altman DG, et al. The revised CONSORT statement for reporting randomized trials: Explanation and elaboration. Ann Intern Med 2001;134:663-94.
11. Brittain E, Lin D. Comparison of intent-to-treat and per protocol results in antibiotic non-inferiority trials. Stat Med 2005;24:1-10.
12. Jones B, Jarvis P, Lewis JA, Ebbutt EF. Trial to assess equivalence: The importance of rigorous methods. BMJ 1996;313:36-9.
13. Dunnet CW, Gent M. Significance testing to establish equivalence between treatments with special to data in the form of 2x2 tables. Stat Med 1996;15:1729-38.
14. Dunnet CW, Gent M. An alternative to the use of two-sided tests in clinical trials. Biometrics 1977;33:593-602.
15. Rothman KJ. Significance questing. Ann Intern Med 1986;105:445-7.