

# Clinical Research Methods

## Non-inferiority clinical trials: What they are and when they should be done

PREM PAIS

### ABSTRACT

Randomized controlled trials (RCTs) are the gold standard design of studies evaluating novel interventions, pharmaceutical and otherwise. The standard RCT design is a 'superiority' study to assess if the tested intervention is better than existing treatments. However, a new intervention that has additional benefits, such as better adverse effect profile, may be useful even if it is not more effective than existing treatments. In such cases, a non-inferiority (NI) design may be preferable. In NI designs, once NI is demonstrated, sequential analysis for superiority may be done, provided it is pre-specified. This article discusses the utility and principles of an NI design. Examples to illustrate the design are taken from published trials of anticoagulation in atrial fibrillation.

Natl Med J India 2026;39:120–3

### INTRODUCTION

Randomized controlled trials (RCTs) are the gold standard studies to prove the efficacy of a novel intervention in clinical medicine. Besides oncology, cardiology is one of the disciplines in which such trials have revolutionized the treatment of disease, especially acute ischaemic syndrome. Consider that until 1961, hospital mortality of ST elevation myocardial infarction was in the region of 30%.<sup>1</sup> Presently, such mortality is under 10%.<sup>2</sup> The concepts of managing an acute myocardial infarction (AMI)—rhythm monitoring in a dedicated coronary care unit, antiplatelet and anticoagulation therapy, thrombolysis and percutaneous coronary intervention (PCI), among others—have brought this about. These interventions have been mainly established as a result of RCTs. Similarly, in atrial fibrillation (AF), the risk of stroke and systemic embolism (SE) has been reduced by over 60% with the introduction of vitamin K antagonists.<sup>3</sup> Stroke prevention in non-valvular AF (NVAf) has continued to evolve more recently with the introduction of newer, safer, and more convenient anticoagulants. Here again, RCTs have played a key role.

Traditionally, RCTs aim to show that a new intervention is superior to no treatment or existing therapy. With medical advances, new treatments usually offer incrementally small but clinically useful benefits over existing treatments. Newer treatments with features other than efficacy may emerge, making them potentially attractive compared to existing treatments. These additional benefits would make them useful even if they were as effective or almost as effective as

existing treatments. It is in such situations that non-inferiority (NI) RCTs are useful.

This article explains the concepts and utility of such a trial design using trials of non-vitamin K oral anticoagulants (NOACs) in NVAf as examples. These trials lend themselves conveniently to illustrate such a trial design. Further, the author has been actively involved in these trials.

### SUPERIORITY RCT DESIGNS

The stroke prevention in AF (SPAF) trial<sup>3</sup> was a typical RCT designed to demonstrate the superiority of a new treatment and compared the use of the vitamin K antagonist, warfarin, with placebo. The results of this trial clearly demonstrated the superiority of warfarin in preventing stroke and systemic embolism—warfarin compared to placebo reduced the risk of stroke by a point estimate of 33%, with a minimum of 15% reduction (relative risk [RR] 0.67; 95% CI 0.27–0.85,  $p=0.01$ ).

### *Null hypothesis in a superiority RCT*

In the analysis of a clinical trial, rather than prove the hypothesis being tested, the statistician sets out to disprove the 'null' hypothesis. If such a null hypothesis can be statistically rejected, the alternative hypothesis stands. The null hypothesis in a superiority trial would be that the new treatment is 'not superior' to the control treatment. If this is disproved, the alternate hypothesis, that the new treatment is 'superior' to control, stands. To disprove the null hypothesis in a superiority trial where the outcome of interest is death, occurrence of disease, or a serious condition, the RR should be less than 1, and the upper limit of the 95% CI should also be within 1, where 1 represents a neutral effect. Further, the  $p$  value would be expected to be  $<0.05$ , indicating a less than 5% chance that the result is due to chance.

In the example from the SPAF trial quoted above, all these criteria are met, and the null hypothesis could be rejected and the alternate hypothesis of superiority established. Thus, in SPAF, warfarin was demonstrated to be superior to placebo in preventing stroke in NVAf. The results of the SPAF trial led to warfarin (or other vitamin K antagonists) becoming the standard of care in stroke prevention in AF. The clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events (CHARISMA) trial<sup>4</sup> offers a contrast. CHARISMA was also a superiority trial testing if a combination of aspirin and clopidogrel is superior to aspirin alone in stable cardiovascular disease (CVD). In CHARISMA, the RR was 0.93 (95% CI 0.83–1.05,  $p=0.22$ ). In this case, while the point estimate (RR) was less than 1, the upper limit for the 95% CI 1.05 was greater than 1, and the  $p$  value did not achieve a level of significance. Thus, in CHARISMA, the null hypothesis could not be rejected. Clopidogrel plus aspirin was not superior to aspirin alone in this population.

Division of Clinical Research and Training, St. John's Research Institute, Bengaluru, Karnataka, India; [pais.prem@gmail.com](mailto:pais.prem@gmail.com)

[To cite: Pais P. Non-inferiority clinical trials: What they are and when they should be done. *Natl Med J India* 2026;39:120–3. DOI: 10.25259/NMJI\_1051\_2024]

© The National Medical Journal of India 2026

## NI DESIGNS

The objective of an NI design in an RCT is to test whether a new intervention or treatment is as good or almost as good (non-inferior) to the standard treatment. The null hypothesis for such a trial would be that the new treatment is inferior to the control treatment, and the alternative hypothesis would be that the new treatment is 'not inferior' to the control treatment (Box 1). If the null hypothesis is rejected, the alternative hypothesis that the new treatment is non-inferior to the control treatment is established. A NI design obviously can only be considered when the new treatment is being compared to an existing standard and effective treatment. It would not make sense to use an NI design when control is a placebo. An advantage of this design is that if NI is established, superiority can be tested for subsequently. This is called sequential analysis, but to be valid, such an analysis should be specified in advance in the study protocol, not post-hoc after the data is available for analysis. It may be noted that such a pre-specified sequential analysis is not considered multiple hypothesis testing and does not result in loss of alpha power of the study.

### *Where would one consider using an NI design in an RCT?*

Consider a situation where a proven and effective medication exists for a potentially severe disease, and is the standard of care. A new molecule becomes available, which potentially has advantages over the existing medication. It may be safer, more convenient to administer, have better pharmacodynamics, and/or cost less. Because of these advantages, the new molecule may be useful even if it is as effective or almost as effective as the standard medication.

How would one design an RCT for the new molecule? Since an effective drug is available, testing against a placebo would be unethical. The new molecule would need to be compared to the existing standard therapy. With a superiority design, if the null hypothesis (Box 1) could not be rejected, the design would not permit statements of it being 'as good' or 'almost as good'. Regulatory authorities would be unlikely to license the new

molecule, and the potential benefits would be lost. This is a situation for the use of an NI design.

A concrete example illustrates this well. Warfarin and vitamin K antagonists (VKAs) are effective in preventing stroke and SE in AF and have been the standard of care until recently. However, their unstable pharmacokinetics/pharmacodynamics (PK/PD) result in significant inter and intra-individual dose variability, necessitating frequent INR monitoring, dietary restrictions, and carry a risk of intracerebral haemorrhage (ICH). NOACS offered some advantages over warfarin—more predictable absorption and PK/PD and blood levels making INR monitoring unnecessary, use of a fixed dose regimen, and possibly less ICH. It seemed that with these potential advantages, even if their efficacy was not superior to warfarin but was as good or even almost as good in preventing stroke/SE in AF, they would be useful. Hence, NI designs were used in an RCT comparing NOACS to warfarin.

Since these molecules were initially focused on more developed countries, testing was mainly done in NVAF. This is the case in both the RELY trial (dabigatran versus warfarin in patients with AF)<sup>5</sup> and ARISTOTLE trial (apixaban versus warfarin in patients with AF).<sup>6</sup> In contrast, the AVERROES (apixaban in patients with AF) trial was a superiority trial.<sup>7</sup> This was done because in AVERROES, apixaban was compared to a placebo in patients with NVAF considered unsuitable for VKA.

### *How would NI be defined?*

When designing an NI trial, an NI margin must be defined in advance. An NI margin is a predetermined margin of difference between the new and standard treatments, which represents how much loss of efficacy one is prepared to accept in the new treatment compared to the control treatment, keeping in mind the potential advantages of the new treatment. To arrive at such a margin, one should first consider the benefit of the existing active control treatment against placebo or other control medication when it was originally tested. The NI margin should preserve a major part of this benefit. Thus, one would have to review the data from RCTs that established the efficacy of the control medicine over placebo, either from a large definitive trial or a meta-analysis that led to the regulatory approval of the control drug. The upper margin of the 95% CI of the control drug against placebo is identified and is termed M1. For setting the NI margin, M2, we should preserve much of this benefit. Thus, for designing an NI trial of a new anticoagulant against warfarin in AF, M1 would be derived from the results of the SPAF trial referred to above in which warfarin was tested against placebo. The results of the SPAF trials showed that warfarin would result in a 33% point-estimate reduction of events compared to placebo, with a minimum of 15% reduction (RR 0.67; 95% CI 0.27–0.85,  $p=0.01$ ). The NI margin should be set to preserve this 15% benefit. For NOAC trials, generally, the NI margin is set as less than 1.46. That is, any result in which the upper confidence limit is within 1.46 will be accepted as non-inferior. A point midway between the 95% CI of the RR reduction was taken as explained in Box 2.

Consider the Rivaroxaban Once Daily Oral Direct Factor Xa Inhibition Compared with Vitamin K Antagonism for Prevention of Stroke and Embolism Trial in Atrial Fibrillation (ROCKET AF) trial.<sup>8</sup> Its primary objective was to compare the effect of rivaroxaban 20 mg once daily with dose-adjusted warfarin in the prevention of stroke and SE in patients with NVAF. An NI design was used. As described above, the NI margin (M2) was

### **Box 1: The null and alternative hypotheses in different clinical trial designs**

#### *Superiority trial*

- NH: New treatment **is not** different (superior) than control treatment
- AH: New treatment **is** different (superior) than control treatment

#### *NI trial*

- NH: New treatment **is inferior** to control treatment
- AH: New treatment **is non-inferior** to control treatment

#### *Sequential analysis*

- NH: New treatment is non-inferior but **not** better than control treatment
- AH: New treatment is non-inferior and **is** better than control treatment

#### *Equivalence trial*

- NH: New treatment **is not** similar to control treatment
- AH: New treatment is similar to control treatment

NH null hypothesis      AH alternative hypothesis

**Box 2: Calculating the non-inferiority margin**

- Warfarin benefit over placebo in definitive trial:
- RR reduction 33%; 95% CI 73%–15%, i.e. minimum of 15% reduction in events versus placebo
- Non-inferiority margin for trial of new non-vitamin K oral anticoagulant should preserve this 15% benefit
- A point midway between the 95% CI of the RR reduction of SPAF is 45.5%
- Hence non-inferiority margin is set at <1.46.

SPAF stroke prevention in atrial fibrillation trial

set at 1.46. Thus, to reject the null hypothesis that rivaroxaban was inferior to warfarin, it was required that the upper limit of the 95% CI should be less than 1.46. A sequential analysis for superiority was planned if NI was achieved. For the latter to be successful, the upper limit of the 95% CI would have to be less than 1. In ROCKET AF, there were 269 primary events in the 7081 patients randomised to rivaroxaban (2.1 events per 100 patient years) and 306 such events in 7090 patients in the warfarin arm (2.4 events per 100 patient years; HR 0.88; 95% CI 0.75–1.03). Since the upper margin of the CI (1.03) is less than 1.46, the null hypothesis of inferiority was rejected, and it was concluded that rivaroxaban was non-inferior to warfarin in this population (p for NI <0.001). However, since this margin exceeded 1.0, sequential analysis could not reject the null hypothesis that rivaroxaban was not superior to warfarin (p for superiority 0.12). The trial concluded that ‘In patients with AF, rivaroxaban was noninferior to warfarin for the prevention of stroke or SE’. Since the trial also showed reduced intracranial and fatal bleeding in the rivaroxaban arm, despite failure to prove superiority, rivaroxaban was approved for the prevention of stroke and SE in NVAF. It may be noted that such trials will have 2 p values—1 for NI and 1 for superiority, which, as in this case, may go in different directions.

The RELY trial<sup>5</sup> illustrates the issue of NI and sequential superiority particularly well (Fig. 1). The trial tested the efficacy of dabigatran versus dose-adjusted warfarin in a population with NVAF. The study design and endpoints were similar to ROCKET AF, with one difference: Two doses of dabigatran were tested. The trial had three arms—one in which patients were treated with dose-adjusted warfarin (control arm) and two intervention arms—one in which patients received dabigatran in a dose of 110 mg twice daily and a second in which a dose of 150 mg twice daily was given. The results of RELY in summary were (i) dabigatran 110 mg arm versus warfarin (HR 0.91; 95% CI 0.74–1.11) and (ii) dabigatran 150 mg arm versus warfarin (HR 0.66; 95% CI 0.54–0.82). As in ROCKET AF, the NI margin was set at 1.46. The results show that dabigatran 110 mg twice daily (upper 95% CI limit 1.11) was non-inferior but not superior to warfarin—p value for NI <0.001; p value for superiority 0.34. On the other hand, the 150 mg dose (upper 95% CI limit 0.82) was both non-inferior and superior to warfarin, p value for NI as well as for superiority <0.001.

Fig. 2 illustrates the results of the RELY trial. In Fig. 2, the upper CI of the 110 mg dose (1.11) lies within the NI margin (1.46) but not within the superiority margin (1.0), indicating NI but not superiority. In the case of the 150 mg dose, the upper margin of the 95% CI (0.82) lies within both the NI and superiority margin, indicating both NI and superiority. However, while the 110 mg

dose was non-inferior but not superior to warfarin it resulted in lower major bleeding (2.71% per year in the dabigatran arm and 3.36% per year in warfarin arm, p=0.003) as well as lower haemorrhagic stroke (0.12% per year in dabigatran arm, 0.38% per year in warfarin arm, p<0.001). For this reason, it is perceived to have an advantage over warfarin, especially in those at increased risk of haemorrhage and in older adults. The 150 mg twice daily dabigatran arm, besides being superior to warfarin in preventing stroke and SE, produced less haemorrhagic stroke (0.10% per year, p<0.001) and a similar risk of major bleeding (0.311% per year, p=0.31) and is considered the standard dose in NVAF.<sup>9</sup> Although not of direct relevance, it is interesting to note that in comparing the 110 mg and 150 mg doses of dabigatran, the higher dose significantly reduced the rate of stroke/SE (p=0.005) but with a strong trend to an increased risk of major bleeding (p=0.052). The rate of death and haemorrhagic stroke (as against all major bleeds) was similar for the two doses.

**NI DESIGN VERSUS EQUIVALENCE DESIGN**

As discussed above, the objective of an NI design is to show that the tested intervention is not significantly worse than the control. Further, the design permits analysis for superiority once NI is established. An equivalence design aims to show that the tested drug is similar to the control in either direction, i.e. the two drugs are ‘equivalent’. In this design, two boundaries of equivalence are preset—an upper and a lower boundary to define equivalence. The design does not permit sequential testing for superiority. This design is not commonly used in clinical trials but may be used to compare the action of two products deemed to be similar. The ASSENT-2 (single-bolus tecteplase compared to front-loaded alteplase in AMI) trial is an example.<sup>10</sup> This trial compared rapid infusion of alteplase with a single bolus of tecteplase in 16 949 patients with AMI. The primary outcome was to show equivalence in all-cause mortality at 30 days between the two arms. Equivalence was defined as 1% absolute or 14% relative

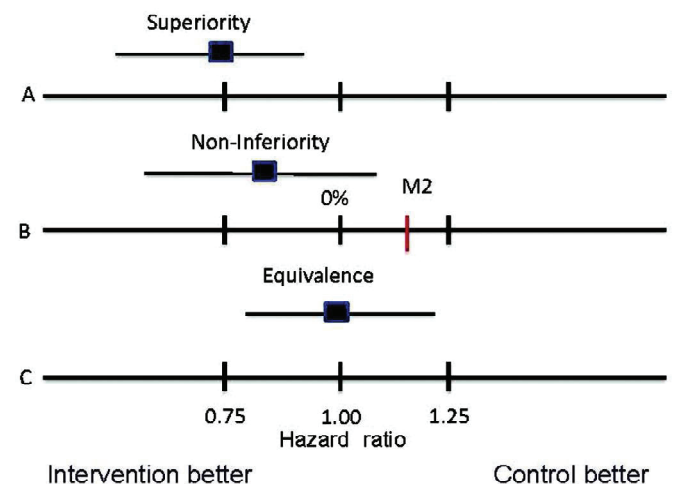


FIG 1. Diagram to illustrate the difference between superiority, non-inferior (NI), and equivalence designs. In superiority trials (A), the point estimate (black square) as well as the 95% CI (lines on either side of the point estimate) should be less than an HR of 1.00, showing clear statistical superiority over control treatment. In a NI trial (B), the point estimate may be on either side of an HR 1.00, but the upper 95% CI should be within the preset NI margin M2 (see text). In an equivalence trial (C), both the point estimate and the 95% CI should mirror the control

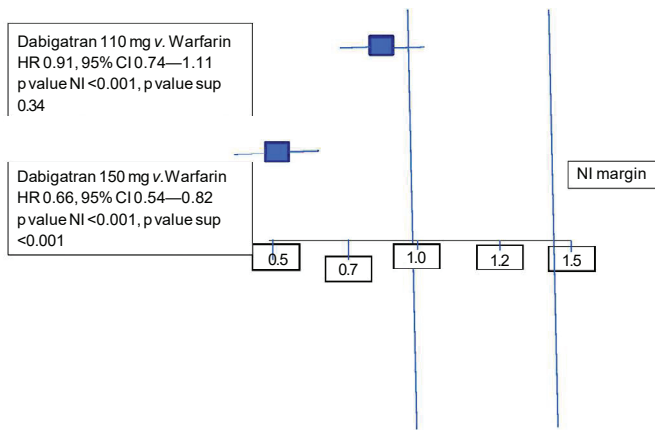


FIG 2. Results of the RELY trial illustrating the concept of non-inferiority (NI) and superiority (sup) in the 2 doses of dabigatran used in the trial versus warfarin HR hazard ratio 95% CI 95% confidence interval

difference in 30-day mortality, one-sided. The 30-day mortality was almost identical in the two arms: 6.18% for tenecteplase and 6.15% for alteplase, and 95% one-sided upper boundaries of the absolute and relative differences in 30-day mortality were 0.61% and 10%, respectively, which were within the pre-specified equivalence criteria, p for equivalence was 0.006. The conclusion was that tenecteplase and alteplase were equivalent for 30-day mortality, but that ease of administration could give tenecteplase the edge.

#### DRAWBACKS OF AN NI DESIGN

An NI design is not without drawbacks. Setting an NI margin is difficult and somewhat subjective. The margin can affect the outcome substantially. The NI margin should be carefully assessed to see whether it is clinically relevant and reasonable. An intention-to-treat analysis, which includes all randomized study participants, even those who have stopped study intervention, is the norm for clinical trials, while a per-protocol analysis, in which only participants who have adhered to the intervention, may be done for safety purposes. Ideally, both analyses should be carried out, and NI should be declared only if both analyses support it.<sup>11</sup> It should also be clarified

again that an NI design would not be appropriate where an intervention is being tested to show it is more effective than existing treatment, and certainly not in a placebo-controlled trial.

#### CONCLUSION

An NI design is chosen when testing a new treatment against an existing effective treatment. The new treatment may have advantages other than greater efficacy, such as lower cost, greater convenience of administration, or greater safety. Use of a superiority design would lead to 'loss' of a useful intervention because it would simply fail to show superiority. On the other hand, an NI design will allow NI to be tested and, if proven, will permit the use of the other advantages of the new treatment. It would also permit testing for superiority if NI were established.

*Conflicts of interest.* None declared

#### REFERENCES

- 1 Nabel EG, Braunwald E. A tale of coronary artery disease and myocardial infarction. *N Engl J Med* 2012;**366**:54–63.
- 2 Xavier D, Pais P, Devereaux PJ, Xie C, Prabhakaran D, Joshi P, *et al.* Treatment and outcomes of acute coronary syndromes in India (CREATE): Prospective analysis of registry data. *Lancet* 2008;**371**:1435–44.
- 3 McBride R for Stroke Prevention in Atrial Fibrillation Investigators. Stroke prevention in atrial fibrillation study. Final results. *Circulation* 1991;**84**:527–39.
- 4 Bhatt DL, Fox KAA, Hacke W, Berger PB, Black HR, Boden WE, *et al.* Clopidogrel and aspirin versus aspirin alone for the prevention of atherothrombotic events. *N Engl J Med* 2006;**354**:1706–17.
- 5 Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom E, Oldgren J, Parekh A, *et al.* Dabigatran versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2009;**361**:1139–51.
- 6 Granger CB, Alexander JH, McMurray JJV, Lopez RD, Hylek EM, Hanna M, *et al.* Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med* 2011;**365**:981–92.
- 7 Connolly S, Eikelboom J, Joyner C, Diener H, Hart R, Golitsyn S, *et al.* Apixaban in patients with atrial fibrillation. *N Engl J Med* 2011;**364**:806–17.
- 8 Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, *et al.* Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. *N Engl J Med* 2011;**365**:883–91.
- 9 Ferreira J, Ferreira D, Viana-Baptista M, Bettencourt P, Cernadas R, Crespo F. Dabigatran for stroke prevention in nonvalvular atrial fibrillation: Answers to challenging "real-world" questions. *Thrombosis* 2012;**2012**:867121.
- 10 Van de Werff F, Adgey J, Ardissino D, Armstrong PW, Aylward P, Barbash G, *et al.* Single-bolus tenecteplase compared with front-loaded alteplase in acute myocardial infarction: The ASSENT-2 double-blind randomised trial. *Lancet* 1999;**354**:716–22.
- 11 Snapinn SM. Noninferiority trials. *Curr Control Trials Cardiovasc Med* 2000;**1**:19–21.