

Title: Challenges in the design and interpretation of non-inferiority trials

Subtitle: Insights from recent stent trials

Brief title: Non-inferiority trial design and interpretation

Word count (from introduction to conclusion including headings, figure/table titles and legends, and references): 5,091

Condensed abstract word count: 76

Key words: Clinical trial, Clinical trial design, Non-inferiority trial, Non-inferiority margin, sample size

Authors: Fernando Macaya MD^a, Nicola Ryan MB, BCh^a, Pablo Salinas MD, PhD^a, Stuart J. Pocock PhD^{b,c}

Authors' affiliations:

a. Department of Interventional Cardiology, Hospital Clínico San Carlos – Universidad Complutense Madrid, Madrid, Spain

b. Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom

c. Centro Nacional de Investigaciones Cardiovasculares Carlos III, Madrid, Spain

Authors' email addresses: fernando@macaya.eu, nicolaryan@gmail.com, salinas.pablo@gmail.com, stuart.pocock@lshtm.ac.uk

Conflict of interest: none

Funding: none

Corresponding author: Dr. Pablo Salinas

Postal address: Hospital Clínico San Carlos. Calle del Profesor Martín Lagos s/n.
28040, Madrid, Spain.

Email: salinas.pablo@gmail.com

[Phone: +34913303283](tel:+34913303283) / [Fax: +34913303091](tel:+34913303091)

CONDENSED ABSTRACT

The non-inferiority design is used extensively in current clinical research, but its complex features may hamper the appropriate interpretation of such trials. Thus, understanding the pillars of non-inferiority design is indispensable. We discuss fundamental concepts regarding the design and interpretation of non-inferiority trials and then explore some common methodological criticism by analysing a sample of contemporary coronary stent trials. Finally, we give an overall perspective in order to enhance the design and conduct of future trials.

ABBREVIATIONS AND ACRONYMS:

NI = Non-inferiority

CI = Confidence interval

CONSORT = Consolidated Standards of Reporting Trials

FDA = Food and Drug Administration

TAVR = Transcatheter aortic valve replacement

SAVR = Surgical aortic valve replacement

RR = Relative risk

OR = Odds ratio

HR = Hazard ratio

MANUSCRIPT

Over the past 15 years there has been an increased use of non-inferiority (NI) designs for randomized controlled trials, especially in cardiology(1,2). This is due in part to the challenge that new treatments have to compete with pre-existing effective standard treatments, making placebo-controlled trials unethical in many situations. NI trials assess the hypothesis that a new treatment is not unacceptably worse, regarding a specific efficacy (or safety) criterion, than a standard treatment, usually an active control.

This design has facilitated the approval of many new drugs and devices proposed as valuable alternatives to standard therapy, given an anticipated ancillary benefit (e.g. fewer side effects, higher drug adherence or lower procedural risk). Despite their common use, these studies remain poorly understood due to their statistical complexity and difficulties of interpretation. Moreover, criticism has been raised concerning their various design limitations(3,4). In this paper, we aim to provide insight regarding the essential concepts required for appropriate interpretation of NI trials, as well as focus on some controversies illustrated with recent examples, especially trials of coronary stents.

PRINCIPLES OF A NON-INFERIORITY TRIAL

Classic superiority clinical trials propose new treatments to replace (previous treatment as comparator) or improve (placebo as comparator) current standard practice by outperforming it in terms of major efficacy and/or safety clinical endpoints. Conversely, a NI trial proposes an alternative treatment to be as good as (not inferior to) the existing standard, usually with an ancillary benefit(5–7). These are not necessarily opposing

facets: for example, transcatheter aortic valve replacement (TAVR) initially emerged as a treatment for aortic stenosis in inoperable patients, replacing medical treatment by demonstrating superiority in clinical outcomes including mortality(8). At the same time, TAVR was also proposed as an alternative to surgical aortic valve replacement (SAVR), the gold standard treatment, offering ancillary benefits (less invasive approach, avoidance of cardiopulmonary bypass and shorter ICU stay) and potentially comparable clinical outcomes. Consequently, TAVR was tested in operable patients with several surgical risk profiles and proved to be not inferior in hard clinical outcomes, including mortality and stroke, to SAVR(9,10). A NI trial may also include the hypothesis of superiority as a primary outcome, once the first hurdle of establishing non-inferiority is achieved e.g liraglutide in the LEADER study(11).

Comparing a new treatment to a recognised standard requires justification on grounds of a compelling mechanism of action and a positive framework of pre-trial evidence. For example, the pivotal NI trial ABSORB III(12) proposed biorresorbable scaffolds (Absorb™, Abbott Vascular) as an alternative to drug-eluting stents, using the rationale of an extremely appealing concept (non-permanent vessel scaffolding) and favourable experiences in small clinical cohorts(13).

In practice, this general model is sometimes adopted for slightly different purposes e.g. aiming for a combination of non-inferior efficacy (NI design for composite ischemic event) and improved safety (superiority design for major bleeding) as in the OASIS- 5 study(14); excluding a safety concern in a treatment with known efficacy (sitagliptin and cardiovascular safety in the TECOS study(15)); making head-to-head comparisons of treatments currently in clinical practice (cryoballoon or radiofrequency ablation for paroxysmal atrial fibrillation in the FIRE AND ICE study(16)); and so-called “me-too”

trials, which introduce similar medications or devices with subtle improvements, but not substantial advantages (coronary drug-eluting stent trials(12,17–24)).

DEFINING A RELEVANT DIFFERENCE: THE NON-INFERIORITY MARGIN

The non-inferiority margin, delta (Δ), is a critical component when considering the definition of “not being worse” in NI trials. This NI 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. RR, OR, HR).

The NI margin is fixed in advance and should be clinically justified. It ideally represents the smallest evidence of inferiority, which if true would mean that the new treatment is unacceptable. In the PARTNER 2 trial comparing TAVR with SAVR, a relative NI margin (Hazard Ratio) of 1.20 was chosen for the composite endpoint of death and disabling stroke. The study showed a HR of 0.89 (95% CI 0.73 to 1.09) for TAVR and therefore the authors stated that TAVR appeared not inferior to the surgical therapy for this pre-specified margin(10).

The magnitude of a NI margin 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. Too conservative (narrow) a margin may lead to a large and unfeasible trial with the risk of inconclusive results for truly non-inferior therapies. On the other hand, too liberal a margin may allow moderately inferior therapies to enter clinical practice based on insufficient evidence, potentially becoming inappropriate new standards for future non-inferiority trials. If the latter occurs repeatedly, a paradox may emerge when the active control becomes no better than placebo (so-called ‘biocreep’ phenomenon(25)). When feasible, the margin is chosen based on the known effect of

the active control versus placebo (e.g. choose half of this as margin). However, in stent/TAVI trials the margin seems a more arbitrary choice influenced by what is a realistic sample size.

In order to preserve the active control effect, it is key to choose the best available standard for the active control as well as selecting a reasonable and justified non-inferiority margin. Choice of too liberal (large) margin and hence too small a trial increases the risk of false claims of non-inferiority. Such a large margin may be convenient for trial size but clinically unacceptable. Both statistical and clinical judgement are required when choosing the non-inferiority margin and several key factors should be taken into consideration (**Table 1**). Detailed technical approaches to appropriately choose the NI margin and caveats to bear in mind are discussed elsewhere(3,5,6,25).

ASSESSING TREATMENTS' DIFFERENCES

From a statistical viewpoint, a NI trial is primarily interested in only one direction of potential treatment difference, namely whether the experimental therapeutic measure is (or is not) worse than the current standard(25). However, these trials can still go on to superiority test (two-sided) once the NI criterion is met. Therefore, several different scenarios may occur each with their own interpretation (**Figure 1**).

For a new treatment to be considered non-inferior (i.e. achieve statistical significance for the test of non-inferiority) the confidence interval of the observed difference with the active control must exclude the NI margin (**Figure 1, situations A, B, C**). If this occurs the authors can reasonably assume that the true effect is as good as the current

standard for the given margin. The possibility of showing that the experimental treatment is significantly better remains a legitimate objective within these trials' designs(3,6): that is, it is legitimate to test superiority in a trial once the non-inferiority criterion are met, as occurs in **Figure 1, situation A**. Similar to a superiority trial where a 'negative' (non-significant) outcome does not mean the treatments are equivalent, it is important to understand that failing to demonstrate non-inferiority in a NI trial may mean the trial is inconclusive as in **Figure 1, situations D and E** where the CI indicates that the data are compatible with both no difference and a difference size delta. This may well occur if the trial is too small. An infrequent situation (**situation F**) may occur if the interval of the difference excludes the NI margin and at the same time does not include the zero difference (i.e. absolute 0 difference or relative risk 1). In this case, the therapy is significantly inferior, but not to the extent of the pre-defined NI margin. Alternatively, the chosen NI margin may have been unrealistically large.

POWERING THE TRIAL TO DETECT DIFFERENCES

The null and alternative hypothesis in NI trials are switched around compared to a classical superiority trial. The trial is powered to be able to reject the null hypothesis that the experimental treatment has inferiority delta compared to the active control i.e. the trial should provide compelling evidence to favour the alternative hypothesis of no true treatment difference.

Power calculations to determine a NI trial size (not presented here) follow similar principles to superiority trials re type I and type II errors, with subtle nuances. For the specific NI margin chosen it is generally assumed that a one-sided 2.5% or two-sided 5% alpha risk (equivalent to consider a one-sided 97.5% CI or two-sided 95% CI

respectively), and preferably a 10% beta risk (90% power) to ensure confidence in establishing non-inferiority if truly the two treatments have identical efficacy. For device trials, there are instances where a 5% one-sided alpha risk has been permitted(26). This practice (with the consequent CI 14% narrower) makes it easier (too easy?) to claim non-inferiority and reduces the required trial size by 18%. It is generally not accepted for drug trials and is becoming less common for device trials. Sometimes trialists opt for 80% power (rather than 90%)(26) since it reduces the required trial size by 25%, though of course this carries a higher risk of failing to demonstrate non-inferiority.

The two other key parameters in trial size calculation are the magnitude of the difference to be detected (NI margin) and the expected event rate in the control group. The latter is important since an overestimation of the event rate can lead to an underpowered trial. This estimation usually comes from previous trials, meta-analyses or registries. Therefore, if the new trial's population differs from the population from which the expected event rate was derived the observed event rate may also differ(3,7).

In cases where the observed event rate is substantially lower than expected, an interim analysis may lead to a decision to increase trial size to regain power, though this is not always achievable. The OASIS-5 trial was designed to assess non-inferiority of fondaparinux compared to enoxaparin in an efficacy composite event (death, myocardial infarction, or refractory ischemia at nine days) in 16,000 patients with acute coronary syndrome. A blinded review of the first 4000 patients indicated that the overall event rate was lower than expected. Therefore, the inclusion criteria were modified so that patients under the age of 60 years were required to have both an elevation of biomarkers and ischemic electrocardiographic changes, and the sample size was increased to 20,000 patients(14).

CHOICE OF AN ABSOLUTE OR RELATIVE NON-INFERIORITY MARGIN

Suppose an absolute difference (% of events) is chosen instead of a risk ratio (RR, OR, HR) as the NI margin. Then if the observed event rate in the control group is lower than expected, the fixed absolute NI margin becomes proportionally wider (more permissive) in respect to the observed event rate in the control group(3,6).

For instance, imagine a trial design anticipated an event rate of 10% in the active control arm (as for the experimental arm) and chose an absolute NI margin of 3.5%. This is equivalent to a relative risk NI margin = $(3.5\% + 10\%) / 10\% = 1.35$; that is a maximum relative increase of 35% of events accepted as non-inferior for the experimental arm.

1000 patients in each arm were recruited to provide at least 85% power. Suppose the trial's observed event rates were 5.5% and 6.0% in the active control and experimental arms respectively, yielding an absolute risk difference 0.5% with two-sided 95% CI (one-sided alpha 2.5%) from -1.5 to 2.5%. This meets the pre-defined non-inferiority criterion (upper CI bound < 3.5%) with a significant p value for non-inferiority ($p < 0.05$).

The problem here, is that the corresponding relative risk increase $6.0\% / 5.5\% = 1.09$ with a 95% CI from 0.76 to 1.56. That is, the upper limit far exceeds the relative margin of 1.35 considered in the design. The statistical test for non-inferiority was 'tricked' by choosing an absolute NI margin in the beginning, that in the end is equivalent to a relative risk margin of 1.64 $[(3.5\% + 5.5\%) / 5.5\%]$ which appears unacceptably large.

Hence, in this hypothetical case, an absolute NI margin in combination with a lower than expected event rates permitted higher relative rate increases to be accepted as non-

inferior. If instead a relative NI margin of 1.35 was chosen, it would have led to the inability to claim non-inferiority. The trial would be declared inconclusive (similar to scenario E of Figure 1).

Using relative indices has the advantage of guarding against unrealistically 'optimistic' claims of non-inferiority if the control group event rate is lower than expected, but in such circumstances a trial will need to be larger to achieve its aims. That is, the power of a NI trial with an appropriate relative risk NI margin requires a certain number of primary events. Hence, if the event rate is lower than expected, the number of patients needs to be increased accordingly.

INTENTION TO TREAT AND PER PROTOCOL ANALYSIS

High quality trial conduct is always sought, but is particularly important in NI trials. Low adherence, crossovers, loss to follow-up and misclassification of endpoints tend to make two treatments' results appear similar and therefore bias towards the non-inferiority hypothesis(7).

In this regard, intention-to-treat (ITT) analysis, standard in superiority trials as the most robust outcome analysis, may facilitate similarity between treatments making it easier to demonstrate non-inferiority. For instance, in a trial with many crossovers the differences are diluted and thus the ITT analysis will show artificially similar outcomes between groups. Conversely, per-protocol (PP) or as-treated (AT) analysis may preserve subtle treatment differences thus reducing the risk of false claims of non-inferiority. Thus, PP/AT analyses are of particular value in the reporting for NI trials, despite their risk of potential bias e.g. compliers may be an unrepresentative subset of patients. Moreover,

ITT preserves the advantages of randomisation and provides a result closer to the overall effect that the treatment would have in a real-life scenario. Ideally, both analysis should be reported in order to evaluate the consistency of the results(3,6). For example, PROCEED II study used a NI design to compare two different methods of preserving human donor hearts (i.e. cold storage versus ex-vivo perfusion). The primary endpoint of patient and graft survival was analysed and reported in ITT, PP and AT populations, demonstrating noninferiority as a consistent finding in all three analyses(27).

OFFICIAL STATEMENTS

Guidelines regarding the conduct and reporting of NI trials have been published by several groups including the Consolidated Standards of Reporting Trials (CONSORT) Statement(28), the FDA(29), and the European Medicines Agency (EMA)(30). All recommend outlining the rationale for choosing a noninferiority design. Other important recommendations from both CONSORT and the FDA include an explanation of how the study hypothesis is incorporated into the design, how the participants' interventions and outcomes were chosen, a description of the statistical methods including how the sample size was calculated, and how the study design affects its interpretation and conclusions(28,29).

The importance of a NI margin's influence on trial outcomes is reflected in all guidelines. CONSORT(28) recommend specifying the rationale for the choice of the NI margin along with whether the margin is based on an absolute or relative scale. The EMA is concerned with both the absolute and relative efficacy of the new treatment, although do not provide specific recommendations(30).

Finally the CONSORT guidelines recommend clarification on whether the results are based on an ITT or PP analysis or both, along with commenting on the stability of the results with respect to the different analysis(28). While the FDA guidelines discuss the advantages and disadvantages of an ITT analysis they do not provide definitive recommendations(29).

CURRENT METHODOLOGICAL CONCERNS

Despite efforts from regulatory agencies and the CONSORT group to guide trialists in adequately conducting and reporting non-inferiority trials, some commonly recognised deficiencies persist(1,31,32), in all fields including cardiology(2).

Arbitrarily large NI margins and/or inflated expected event rates have previously been outlined as sources of bias towards potentially false claims of non-inferiority of a new treatment(3,4,6). Both parameters critically determine sample size and therefore study costs and hence it is tempting for trialists to inflate them in order to reduce the sample size. Absence of justification in the selection of the NI margin has also been extensively noted(1,2,28).

A noteworthy use of NI design concerns comparison of alternative stents for percutaneous coronary interventions. Following the introduction of drug-eluting stents, rates of restenosis and clinical events have decreased considerably, and thus new stents are unlikely to show superiority(33). We now perform a systematic analysis of the reporting of NI trials comparing new generation drug eluting stents (DES) to second generation DES.

AN EXAMPLE OF RECENT CORONARY STENT TRIALS

Based on a Medline search, we identified 9 NI trials comparing new generation stents to second generation stents, published in high impact journals during 2010 to 2015 (12,17–24). All trials had similar target populations and a common primary endpoint (with subtle variations among them): Target Lesion Failure (TLF), defined as a composite of cardiac death, target vessel myocardial infarction and target lesion revascularization.

An overview of these studies (**Table 2**) reveals the following: most (7/9) trials reported the source justifying the expected control TLF rate (asterisks). In all studies, except one, the observed event rate was lower than expected, markedly so in some trials. Note all trials had an absolute risk difference as the chosen NI margin, expressed in terms of absolute percentage of events. Remarkably, the NI margins varied from 2.5% to 8.6% (median 3.6%) and only two trials justified the selection of this value (asterisks).

Moreover, all trials but ABSORB III assumed a 5% one-sided alpha risk for testing non-inferiority.

As mentioned above, one consequence of using an absolute NI margin and having a lower event rate than expected is that it permits higher relative rate increases to be accepted as non-inferior. We aimed to demonstrate this by calculating the corresponding relative risk NI margins used in these studies [$\text{relative NI margin} = (\text{event rate} + \text{absolute NI margin}) / \text{event rate}$]. When plotting the observed absolute risk difference with CI (two-sided 90 or 95% according to their design) along with the pre-defined absolute NI margins (**Figure 2, left panel**) all studies show non-inferiority for the new treatment. However, when plotting the relative risk between groups with CI's and the corresponding calculated relative NI margin (**Figure 2, right panel**) used in the design,

the trial findings become more inconclusive; only 4 out of 9 trials consistently demonstrated non-inferiority using this relative risk criterion.

The previous analyses were done on an intention-to-treat basis, which was the primary reported outcome in all the trials. **Table 3** shows the same analysis performed in a per protocol or as-treated basis (reported by 7 of the 9 trials and showing similar conclusions) and performed assuming the recommended one-sided alpha 2.5% versus 5% assumed in the majority of the studies.

This review illustrates how fixed absolute NI margins in combination with substantial (apparently unexpected) decreases in event rates can artificially enhance claims of non-inferiority. Furthermore, large and potentially unjustified NI margins (either absolute or relative) facilitate this bias. The variability evident in the choice of the NI margins across trials reflects a lack of consensus which needs clearer resolution. Moreover, alpha risk assumed in most of these studies does not strictly comply with current recommendations.

A CONSTRUCTIVE APPRAISAL FOR FUTURE NON-INFERIORITY TRIALS

As previously discussed, too wide a NI margin translates into excessive tolerability for the experimental treatment, whereas too narrow a NI margin may prove too tough a hurdle prohibiting some truly non-inferior therapies from entering clinical practice. The magnitude of the NI margin critically determines the size of the trial; thus, its choice requires a realistic balancing of scientific goals with an achievable sample size.

Given the importance of the pre-determined NI margin, careful consideration should be given to its selection: this may encompass expert trialists actively interacting with

regulatory agencies. Likewise, the rationale for choosing this parameter needs to be clearly documented in the trial protocol and properly reported in the eventual publication.

The choice of a relative or absolute margin remains a topic of debate: to date, no consensus has been reached amongst experts or agencies. Similar to too narrow a NI margin making a trial unfeasible, enforced use of relative scales for the NI margin may restrict some new technologies entering the market. Having said this, the data presented from the above stent trials should raise concern about the deliberate use of absolute margins and its consequences. Active research areas which repeatedly use non-inferiority designs, such as coronary stents, require a better consensus regarding the consistent choice of non-inferiority margins.

In contrast, non-inferiority trials evaluating drugs tend to use a relative rather than absolute risk when choosing a NI margin. A controversy over the potential risk of myocardial infarction in diabetic patients taking rosiglitazone (eventually not confirmed) led to the FDA and EMA mandating cardiovascular safety outcomes for all diabetic medications. Following this mandate, in 2007, numerous non-inferiority trials have been carried out in this field. Of note these trials have a common approach (as recommended by FDA guidance) all using a relative risk of 1.3(11,15).

Thankfully nowadays we observe relatively low primary event rates with most drug and device trials in cardiology. However, this provides a real challenge for non-inferiority trials. The anticipated event rate in a trial's control arm should be sensible and justified, based on the most recent, contemporaneous and comparable patient cohort i.e. meta-analysis, trial or registries. The trialists should explain their choice of event rate in the trial protocol.

Despite the best efforts of trialists, the event rate may be difficult to predict and is often lower than expected. To tackle this issue which occurs commonly in the field of coronary stents, expected event rates should perhaps be adjusted downwards in advance. Alternatively, to avoid problems of unexpected lower event rates occurring, one should use blinded interim results to gain insight as to whether the trial size should be increased or eligibility shifted to a higher risk population. Another possible solution is to recommend in advance an event-driven follow-up duration in order to achieve the desirable event rate with a fixed number of patients.

The drawback of adopting relative NI margins and adjusting the trial's design for eventual event rate decreases, is that the target trial size may not be attainable. When evaluating the viability of a NI trial, the investigator should reflect on the need for a NI design and what does the new treatment add to our knowledge and patient benefit.

Several late generation stents developed over the past decade have for the most part been assessed in non-inferiority trials compared with Xience V. Despite the volume of trials and financial investment, we have not witnessed a substantial clinical improvement with these later generation stents. Nevertheless, successive subtle technological improvements in these devices have led to highly developed platforms, enhancing the feasibility of percutaneous coronary revascularization and providing treatment to a wider patient population.

Moreover, as a scientific community we must reflect on our use (or abuse) of the NI design in the field of coronary stents. Although some consensus has been reached in this field (34), a consistent choice of NI margin should be considered to ensure preservation of the original benefit gained in the predecessor superiority trials. Furthermore, special caution is required when a new device or therapy is proposed as a novel concept based

upon a potential ancillary benefit (e.g. biorresorbable scaffolds). In these cases, the pivotal patient-oriented trials may prove inconclusive (or even controversial).

CONCLUSIONS

As the scientific community becomes more acquainted with cautiously interpreting results based on p values(35,36), readers should be aware that the conclusions of a non-inferiority trial do not rely entirely on a test for non-inferiority or an upper confidence interval bound exceeding a limit. Careful interpretation of the absolute and relative magnitudes of difference between treatments, clinical relevance of events and supplementary benefits of the tested therapy should be taken into account to draw meaningful conclusions from each non-inferiority trial (**Central illustration**).

Non-inferiority trials can be particularly sensitive to bias due to deficiencies in their design and execution. As discussed above it is important to ensure compliance with the correct methodology for their conduct. However, with regard to certain issues, absolute or relative NI margins, and NI magnitude selection, a clearer consensus would be beneficial.

Overall, non-inferiority designs have an important role in cardiovascular therapeutics, and especially in evaluating new coronary stents. We hope our methodological perception illustrated by recent examples, are of value in enhancing the statistical/scientific quality of future non-inferiority trials.

ACKNOWLEDGEMENTS

The authors thank Alfonso de Hoyos for his help creating the forest plots. The first author wishes to acknowledge the Department of Preventive Medicine and Public Health, Facultad de Medicina - Clínica Universidad de Navarra, who have helped him and many other physicians gain a deeper knowledge and understanding of statistics.

REFERENCES:

1. Gopal AD, Desai NR, Tse T, Ross JS. Reporting of noninferiority trials in ClinicalTrials.gov and corresponding publications. *JAMA*. 2015;313:1163–1165. (letter)
2. Vaduganathan M, Patel RB, Samman-Tahhan A et al. Cardiovascular clinical trials with noninferiority or equivalence designs from 2001 to 2012. *Int J Cardiol*. 2016;214:16–8. (letter)
3. Head SJ, Kaul S, Bogers AJJC, Kappetein AP. Non-inferiority study design: lessons to be learned from cardiovascular trials. *Eur Heart J*. 2012;33:1318–24.
4. Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet*. 2007;370:1875.
5. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med*. 2006;145:62–69.
6. Pocock SJ, Clayton TC, Stone GW. Challenging Issues in Clinical Trial Design. *J Am Coll Cardiol*. 2015;66:2886–98.
7. Collier TJ. Tools & Techniques - Statistics: Non-inferiority trials. *EuroIntervention*. 2014;10:526–7.
8. Leon MB, Smith CR, Mack M et al. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med*. 2010;363:1597–607.
9. Smith CR, Leon MB, Mack MJ et al. Transcatheter versus surgical aortic-valve replacement in high-risk patients. *N Engl J Med*. 2011;364:2187–2198.
10. Leon MB, Smith CR, Mack MJ et al. Transcatheter or Surgical Aortic-Valve Replacement in Intermediate-Risk Patients. *N Engl J Med*. 2016;374:1609–20.
11. Marso SP, Daniels GH, Brown-Frandsen K et al. Liraglutide and Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med*. 2016;375:311–22.
12. Ellis SG, Kereiakes DJ, Metzger DC et al. Everolimus-Eluting Bioresorbable Scaffolds for Coronary Artery Disease. *N Engl J Med*. 2015;373:1905–15.
13. Serruys PW, Onuma Y, Ormiston JA et al. Evaluation of the second generation of a bioresorbable everolimus drug-eluting vascular scaffold for treatment of de novo coronary artery stenosis: six-month clinical and imaging outcomes. *Circulation*. 2010;122:2301–12.
14. Yusuf S, Mehta SR, Chrolavicius S et al. Comparison of fondaparinux and enoxaparin in acute coronary syndromes. *N Engl J Med*. 2006;354:1464–76.
15. Green JB, Bethel MA, Armstrong PW et al. Effect of Sitagliptin on Cardiovascular Outcomes in Type 2 Diabetes. *N Engl J Med*. 2015;373:232–42.

16. Kuck K-H, Frnkranz A, Chun KRJ et al. Cryoballoon or radiofrequency ablation for symptomatic paroxysmal atrial fibrillation: reintervention, rehospitalization, and quality-of-life outcomes in the FIRE AND ICE trial. *Eur Heart J*. 2016;37:2858–65.
17. Kimura T, Kozuma K, Tanabe K et al. A randomized trial evaluating everolimus-eluting Absorb bioresorbable scaffolds vs. everolimus-eluting metallic stents in patients with coronary artery disease: ABSORB Japan. *Eur Heart J*. 2015;36:3332–42.
18. Raungaard B, Jensen LO, Tilsted H-H et al. Zotarolimus-eluting durable-polymer-coated stent versus a biolimus-eluting biodegradable-polymer-coated stent in unselected patients undergoing percutaneous coronary intervention (SORT OUT VI): a randomised non-inferiority trial. *Lancet*. 2015;385:1527–1535.
19. von Birgelen C, Sen H, Lam MK et al. Third-generation zotarolimus-eluting and everolimus-eluting stents in all-comer patients requiring a percutaneous coronary intervention (DUTCH PEERS): a randomised, single-blind, multicentre, non-inferiority trial. *Lancet*. 2014;383:413–423.
20. Pilgrim T, Heg D, Roffi M et al. Ultrathin strut biodegradable polymer sirolimus-eluting stent versus durable polymer everolimus-eluting stent for percutaneous coronary revascularisation (BIOSCIENCE): a randomised, single-blind, non-inferiority trial. *Lancet*. 2014;384:2111–22.
21. Saito S, Valdes-Chavarri M, Richardt G et al. A randomized, prospective, intercontinental evaluation of a bioresorbable polymer sirolimus-eluting coronary stent system: the CENTURY II trial. *Eur Heart J*. 2014;35:2021–31.
22. Smits PC, Hofma S, Togni M et al. Abluminal biodegradable polymer biolimus-eluting stent versus durable polymer everolimus-eluting stent (COMPARE II): a randomised, controlled, non-inferiority trial. *Lancet*. 2013;381:651–660.
23. Stone GW, Teirstein PS, Meredith IT et al. A Prospective, Randomized Evaluation of a Novel Everolimus-Eluting Coronary Stent. *J Am Coll Cardiol*. 2011;57:1700–8.
24. Serruys PW, Silber S, Garg S et al. Comparison of zotarolimus-eluting and everolimus-eluting coronary stents. *N Engl J Med*. 2010;363:136–146.
25. D’Agostino RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Stat Med*. 2003;22:169–86.
26. Adams DH, Popma JJ, Reardon MJ et al. Transcatheter Aortic-Valve Replacement with a Self-Expanding Prosthesis. *N Engl J Med*. 2014;370:1790–8.
27. Ardehali A, Esmailian F, Deng M et al. Ex-vivo perfusion of donor hearts for human heart transplantation (PROCEED II): a prospective, open-label, multicentre, randomised non-inferiority trial. *Lancet*. 2015;385:2577–2584.

28. Piaggio G, Elbourne DR, Pocock SJ, Evans SJ, Altman DG, for the CONSORT group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *Jama*. 2012;308(24):2594–2604.
29. Food and Drug Administration. Non-Inferiority Clinical Trials to Establish Effectiveness. Guidance for Industry [Internet]. November 2016. Available from: <https://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf> Accessed May 20, 2017.
30. European Medicines Agency. Guideline on the choice of the non-inferiority margin [Internet]. July 2005. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf Accessed May 20, 2017.
31. Le Henanff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA*. 2006;295:1147–1151.
32. Schiller P, Burchardi N, Niestroj M, Kieser M. Quality of reporting of clinical non-inferiority and equivalence randomised trials-update and extension. *Trials*. 2012;13:214.
33. Byrne RA, Kastrati A. Drug-eluting stent trials: too much non-inferiority, too little progress? *Lancet*. 2014;383:386–388.
34. Byrne RA, Serruys PW, Baumbach A et al. Report of a European Society of Cardiology-European Association of Percutaneous Cardiovascular Interventions task force on the evaluation of coronary stents in Europe: executive summary. *Eur Heart J*. 2015;36:2608–20.
35. Pocock SJ, Stone GW. The Primary Outcome Is Positive — Is That Good Enough? *N Engl J Med*. 2016;375:971–9.
36. Pocock SJ, Stone GW. The Primary Outcome Fails — What Next? *N Engl J Med*. 2016;375:861–70.

FIGURE 1: Forest plot with differences and two-sided CI's showing eight hypothetical outcomes from a non-inferiority trial. Thick dashed vertical line indicates zero difference, neutrality i.e. absolute difference equal to 0 or relative difference (RR, OR, HR) equal to 1. Short dashed vertical line indicates the established NI margin.

FIGURE 2: Forest plot of CI's for absolute differences and relative risks from 9 stent trials. Inverted triangle represents the pre-specified absolute and relative NI margins.

CENTRAL ILLUSTRATION. Challenges in the design and interpretation of non-inferiority trials.