

REVIEW TOPIC OF THE WEEK

Challenges in the Design and Interpretation of Noninferiority Trials

Insights From Recent Stent Trials



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

ABSTRACT

The noninferiority 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 noninferiority design is indispensable. The authors discuss fundamental concepts regarding the design and interpretation of noninferiority trials and then explore some common methodological criticism by analyzing a sample of contemporary coronary stent trials. Finally, the authors give an overall perspective to enhance the design and conduct of future trials. (J Am Coll Cardiol 2017;70:894-903)
© 2017 by the American College of Cardiology Foundation.

Over the past 15 years there has been an increased use of noninferiority (NI) designs for randomized controlled trials, especially in cardiology (1,2). This is due, in part, to the challenge that new treatments must 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 AN NI 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, an 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: using the example of the PARTNER (Placement of Aortic Transcatheter Valves) trials, 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,



Listen to this manuscript's audio summary by JACC Editor-in-Chief Dr. Valentin Fuster.



From the ^aDepartment of Interventional Cardiology, Hospital Clínico San Carlos—Universidad Complutense Madrid, Madrid, Spain; ^bDepartment of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom; and the ^cCentro Nacional de Investigaciones Cardiovasculares Carlos III, Madrid, Spain. The authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Manuscript received May 22, 2017; accepted June 12, 2017.

avoidance of cardiopulmonary bypass, and shorter intensive care unit stay) and potentially comparable clinical outcomes. Consequently, TAVR was tested in operable patients with several surgical risk profiles and proved to be noninferior in hard clinical outcomes, including mortality and stroke, to SAVR (9,10). An NI trial may also include the hypothesis of superiority as a primary outcome, after the first hurdle of establishing NI is achieved (e.g., liraglutide in the LEADER [Liraglutide Effect and Action in Diabetes: Evaluation of Cardiovascular Outcome Results] study) (11).

Comparing a new treatment to a recognized 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 bioresorbable scaffolds (Absorb, Abbott Vascular, Santa Clara, California) as an alternative to drug-eluting stents (DES), using the rationale of an extremely appealing concept (non-permanent vessel scaffolding) and favorable experiences in small clinical cohorts (13).

In practice, this general model is sometimes adopted for slightly different purposes: for example, aiming for a combination of noninferior efficacy (NI design for composite ischemic event) and improved safety (superiority design for major bleeding), as in the OASIS-5 (Organization to Assess Strategies in Acute Ischemic Syndromes-5, Comparison of Fondaparinux and Enoxaparin in Acute Coronary Syndromes) study (14); excluding a safety concern in a treatment with known efficacy (sitagliptin and cardiovascular safety in the TECOS [Sitagliptin Cardiovascular Outcomes Study (MK-0431-082)] 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 [Comparative Study of Two Ablation Procedures in Patients With Atrial Fibrillation] study) (16); and so-called “me-too” trials, which introduce similar medications or devices with subtle improvements, but not substantial advantages (coronary DES trials [12,17–24]).

DEFINING A RELEVANT DIFFERENCE: THE NI MARGIN

The NI 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., relative risk [RR], odds ratio [OR], or hazard ratio [HR]). The NI margin is fixed in advance and should be clinically justified. It ideally represents the smallest evidence

of inferiority that, if true, would mean the new treatment is unacceptable. In the PARTNER 2 trial comparing TAVR with SAVR, a relative NI margin (HR) of 1.20 was chosen for the composite endpoint of death and disabling stroke. The study showed an HR of 0.89 (95% confidence interval [CI]: 0.73 to 1.09) for TAVR, and therefore, the authors stated that TAVR appeared to be noninferior to the surgical therapy for this pre-specified margin (10).

The magnitude of an 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 noninferior therapies. However, too liberal a margin may allow moderately inferior therapies to enter clinical practice based on insufficient evidence, potentially becoming inappropriate new standards for future noninferiority trials. If the latter occurs repeatedly, a paradox may emerge where the active control becomes no better than placebo (so-called biocrep phenomenon [25]). When feasible, the margin is chosen based on the known effect of the active control versus placebo (e.g., choose one-half of this as a margin). However, in stent/TAVR trials, the margin seems to be a more arbitrary choice influenced by what is a realistic sample size.

To preserve the active control effect, it is key to choose the best available standard for the active control as well as to select a reasonable and justified non-inferiority margin. Choice of too liberal (large) a margin, and hence too small of a trial, increases the risk of false claims of noninferiority. Such a large margin may be convenient for trial size, but is clinically unacceptable. Both statistical and clinical judgment are required when choosing the noninferiority 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 TREATMENT DIFFERENCES

From a statistical viewpoint, an NI trial is primarily interested in only 1 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 a superiority test (2-sided) after the NI criterion is met. Therefore, several different scenarios may occur, each with its own interpretation (Figure 1).

ABBREVIATIONS AND ACRONYMS

CI	= confidence interval
CONSORT	= Consolidated Standards of Reporting Trials
FDA	= U.S. Food and Drug Administration
HR	= hazard ratio
NI	= noninferiority
OR	= odds ratio
RR	= relative risk
SAVR	= surgical aortic valve replacement
TAVR	= transcatheter aortic valve replacement

TABLE 1 Key Points to Bear in Mind When Choosing an NI Margin

1. Accurate estimation of the efficacy of the active control (over placebo or previous controls)
2. Nature and clinical relevance of events (misleading composite outcomes)
3. Implications of choosing an absolute or relative difference
4. Weight of the new treatment's ancillary benefit
5. Achievable sample size

NI = noninferiority.

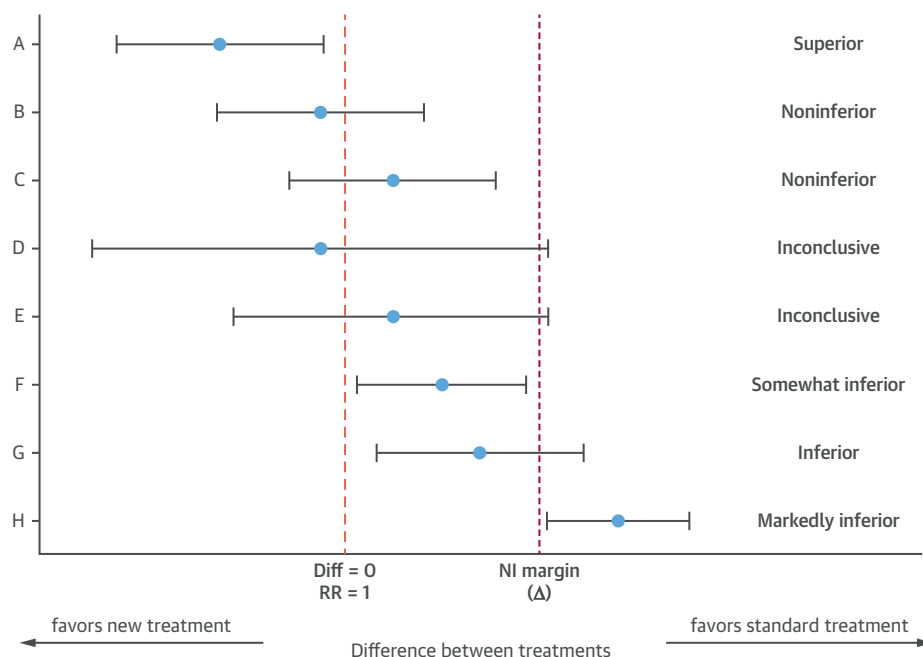
For a new treatment to be considered noninferior (i.e., achieve statistical significance for the test of noninferiority) the CI of the observed difference with the active control must exclude the NI margin (Figure 1, situations A to C). If this occurs, the authors can reasonably assume that the true 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 after the noninferiority criteria are met, as occurs in Figure 1, situation A. Similar to a superiority trial, where a "negative" (nonsignificant) outcome does not mean

the treatments are equivalent, it is important to understand that failing to demonstrate noninferiority in an NI trial may mean the trial is inconclusive; this is the case 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 (Figure 1, 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 RR: 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 reversed 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 favor the

FIGURE 1 Forest Plot With Differences and 2-Sided CIs Showing 8 Hypothetical Outcomes From an NI Trial



Orange dashed vertical line indicates zero difference, neutrality (i.e., absolute difference equal to 0 or relative difference [RR, odds ratio, hazard ratio] equal to 1). **Red dashed vertical line** indicates the established NI margin. CI = confidence interval; diff = difference; NI = noninferiority; RR = relative risk.

alternative hypothesis of no true treatment difference).

Power calculations to determine an NI trial size (not presented here) follow similar principles to superiority trials with regards to type I and II errors, with subtle nuances. For the specific NI margin chosen, it is generally assumed that a 1-sided 2.5% or 2-sided 5% alpha risk (equivalent to a 1-sided 97.5% CI or 2-sided 95% CI, respectively), and preferably a 10% beta risk (90% power), ensures confidence in establishing noninferiority if the 2 treatments have truly identical efficacy. For device trials, there are instances where a 5% 1-sided alpha risk has been permitted (26). This practice (with the consequent 14% narrower CI) makes it easier (too easy?) to claim NI 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), as it reduces the required trial size by 25%, although, of course, this carries a higher risk of failing to demonstrate noninferiority.

The 2 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, because 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 study population differs from the study 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, although this is not always achievable. The OASIS-5 trial was designed to assess the NI of fondaparinux compared with enoxaparin in an efficacy composite event (death, myocardial infarction, or refractory ischemia at 9 days) in 16,000 patients with acute coronary syndrome. A blinded review of the first 4,000 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 NI MARGIN

Suppose an absolute difference (percentage of events) is chosen instead of a risk ratio (RR, OR, 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) with respect to the observed event rate in the control group (3,6).

For instance, imagine that 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 of $(3.5\% + 10\%)/10\% = 1.35$; that is, a maximum relative increase of 35% of events accepted as noninferior for the experimental arm.

A total of 1,000 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 of 0.5% with a 2-sided 95% CI (1-sided alpha 2.5%) from -1.5% to 2.5%. This meets the pre-defined NI criterion (upper CI bound <3.5%) with a significant p value for NI ($p < 0.05$).

The problem here is that the corresponding relative risk increased to $6.0\%/5.5\% = 1.09$, 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 NI 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 to be unacceptably large.

Hence, in this hypothetical case, an absolute NI margin in combination with lower than expected event rates permitted higher relative rate increases to be accepted as noninferior. If, instead, a relative NI margin of 1.35 was chosen, it would have led to the inability to claim noninferiority. The trial would be declared inconclusive (similar to Figure 1, scenario E).

Using relative indexes has the advantage of guarding against unrealistically optimistic claims of NI 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 an 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 it is particularly important in NI trials. Low adherence, crossovers, loss to follow-up, and misclassification of endpoints tend to make 2 treatments' results appear similar, and therefore, bias toward the noninferiority 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 NI. 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 NI. Thus, PP/AT analyses are of particular value in the reporting for NI trials, despite their risk of potential bias (compliers may be an unrepresentative subset of patients). On the other hand, ITT preserves the advantages of randomization and provides a result closer to the overall effect that the treatment would have in a real-life scenario. Ideally, both analyses should be reported to evaluate the consistency of the results (3,6). For example, the PROCEED II (Randomized Study of Organ Care System Cardiac for Preservation of Donated Hearts for Eventual Transplantation) study used an NI design to compare 2 different methods of preserving human donor hearts (i.e., cold storage vs. *ex vivo* perfusion). The primary endpoint of patient and graft survival was analyzed and reported in ITT, PP, and AT populations, demonstrating NI as a consistent finding in all 3 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 U.S. Food and Drug Administration (FDA) (29), and the European Medicines Agency (EMA) (30). All recommend outlining the rationale for choosing an NI design. Other important recommendations from both CONSORT and the FDA include explanations of how the study hypothesis is incorporated into the design and how the participants' interventions and outcomes were chosen; a description of the statistical methods, including how the sample size was calculated; and an explanation of how the study design affects its interpretation and conclusions (28,29).

The importance of an NI margin's influence on trial outcomes is reflected in all guidelines. CONSORT (28) recommends 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 it does not provide specific recommendations (30).

Finally, the CONSORT guidelines recommend clarification on whether the results are based on an ITT analysis, PP analysis, or both, along with commenting on the stability of the results with respect to the different analysis (28). Although 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 NI trials, some commonly recognized 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 toward potentially false claims of NI 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 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 DES, 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- to second-generation DES.

AN EXAMPLE OF RECENT CORONARY STENT TRIALS

Based on a Medline search, we identified 9 NI trials comparing new- to second-generation stents, published in high-impact journals from 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, 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 of 9) trials reported the source justifying the expected target lesion failure rate (Table 2, asterisks). In all studies except 1, the observed event rate was lower than expected, markedly so in some trials. Note that 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 2 trials justified the selection

TABLE 2 Summary of Key Findings From 9 NI Trials of New Stent Versus Second-Generation Stent Comparator

Study Name	% Expected Event Rate	% NI Margin (RR)	Beta	Alpha (1-Sided)	Control Arm			Experimental Arm		
					Stent	n	% Observed Event Rate	Stent	n	% Observed Event Rate
RESOLUTE-AC	8.0	3.5 (1.44)*	0.10	0.050	X	1,126	8.3	RE	1,119	8.2
PLATINUM	5.5	3.5 (1.64)*	0.11	0.050	X/P	737	3.1	P	742	3.5
COMPARE II	9.5	4.0 (1.42)	0.10	0.050	X/P	912	4.8	N	1,795	5.2
CENTURY II	10.0	5.5 (1.55)*	0.10	0.050	X	557	5.0	U	562	4.6
BIOSCIENCE	8.0	3.5 (1.44)*	0.20	0.050	X	1,056	6.6	O	1,063	6.5
DUTCH PEERS	10.0	3.6 (1.36)*	0.20	0.050	P	905	5.2	RI	905	6.1
SORT OUT VI	6.5	2.5 (1.38)*	0.20	0.050	RI	1,497	5.0	B	1,502	5.3
ABSORB III	7.0*	4.5 (1.64)	0.04	0.025	X	677	6.1	A	1,313	7.8
ABSORB Japan	9.0*	8.6 (1.96)*	0.10	0.050	X	133	3.8	A	265	4.2

Relative risk margins (RRs) have been calculated. *Reference provided for value.

A = Absorb (Abbott Vascular, Santa Clara, California); ABSORB III = A Clinical Evaluation of Absorb BVS, the Everolimus Eluting Bioresorbable Vascular Scaffold in the Treatment of Subjects With de Novo Native Coronary Artery Lesions; ABSORB Japan = A Clinical Evaluation of AVJ-301 (Absorb BVS), the Everolimus Eluting Bioresorbable Vascular Scaffold in the Treatment of Subjects With de Novo Native Coronary Artery Lesions in Japanese Population; B = BioMatrix Flex stent (Biosensors Interventional Technologies, Singapore); BIOSCIENCE = A Randomized Comparison of a Sirolimus-eluting Stent With Biodegradable Polymer Versus an Everolimus-eluting Stent With a Durable Polymer for Percutaneous Coronary Revascularization; CENTURY II = Clinical Evaluation of New Terumo Drug-Eluting Coronary Stent System in the Treatment of Patients with Coronary Artery Disease; COMPARE II = Comparison of the Everolimus Eluting (XIENCE-V, XIENCE-Prime or PROMUS Stent) With the Biolimus A9 Eluting NOBORI Stent in All-comers: a Randomized Open Label Study; DUTCH PEERS = Durable polymer-based sTent Challenge of Promus Element versus Resolute Integrity in an all comers population; N = Nobori (Terumo Corporation, Tokyo, Japan); NI = noninferiority; O = Orsiro (Biotronik AG, Bülach, Switzerland); P = Promus (Boston Scientific, Marlborough, Massachusetts); PE = Promus Element (Boston Scientific); PLATINUM = A Prospective, Randomized, Multicenter Trial to Assess an Everolimus-Eluting Coronary Stent System (PROMUS Element) for the Treatment of up to Two De Novo Coronary Artery Lesions; RE = Resolute Endeavor (Medtronic Cardiovascular, Santa Rosa, California); RESOLUTE-AC = RESOLUTE-III All-comers Trial: A Randomized Comparison of a Zotarolimus-Eluting Stent With an Everolimus-Eluting Stent for Percutaneous Coronary Intervention; RI = Resolute Integrity (Medtronic); SORT OUT VI = Randomized Clinical Comparison of Biomatrix Flex and Resolute Integrity Coronary Stents in Non-selected Patients With Ischemic Heart Disease; U = Ultimaster (Terumo); X = Xience (Abbott Vascular).

of this value (Table 2, asterisks). Moreover, all trials except ABSORB III assumed a 5% 1-sided alpha risk for testing NI.

As mentioned earlier, 1 consequence of using an absolute NI margin and having a lower than expected event rate is that it permits higher relative rate increases to be accepted as noninferior. We aimed to demonstrate this by calculating the corresponding relative risk NI margins used in these studies (relative NI margin = [expected event rate + absolute NI margin]/expected event rate). When plotting the observed absolute risk difference with CIs (2-sided 90% or 95%, according to their design) along with the pre-defined absolute NI margins (Figure 2, left) all studies show noninferiority for the new treatment. However, when plotting the relative risk between groups with CIs and the corresponding calculated relative NI margin (Figure 2, right) used in the design, the trial findings become more inconclusive: only 4 of 9 trials consistently demonstrated noninferiority using this relative risk criterion.

The previous analyses were done on an ITT basis, which was the primary reported outcome in all trials. Table 3 shows the same analysis performed in a PP or AT basis (reported by 7 of the 9 trials and showing similar conclusions) and performed assuming the recommended 1-sided alpha of 2.5% versus the 5% assumed in the majority of the studies.

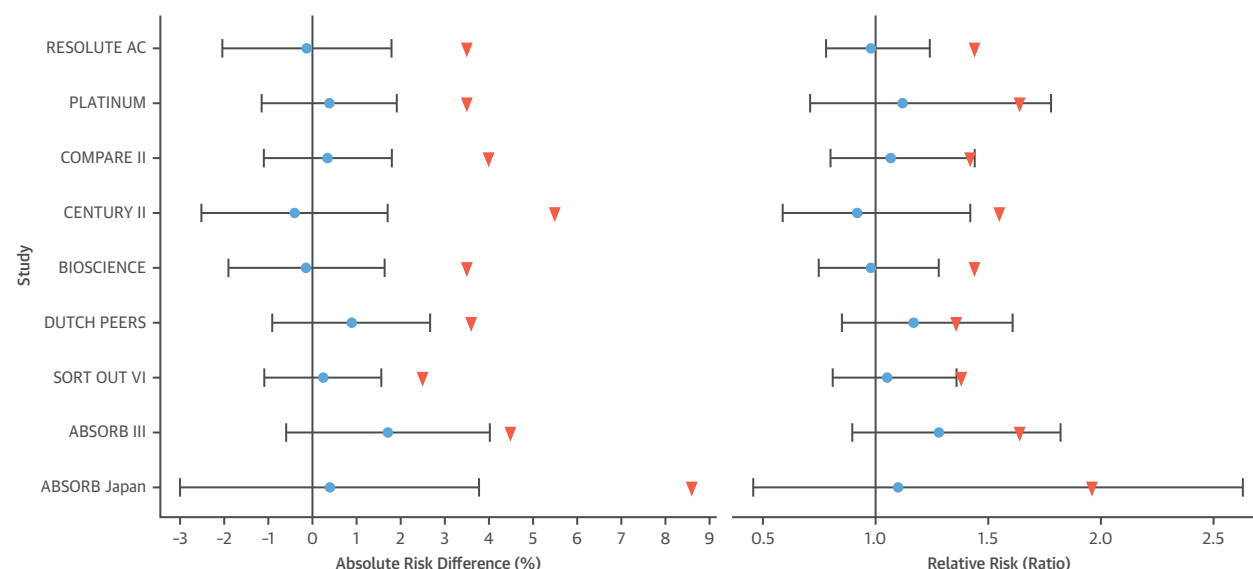
This review illustrates how fixed absolute NI margins in combination with substantial (apparently unexpected) decreases in event rates can artificially enhance claims of NI. Furthermore, large and potentially unjustified NI margins (either absolute or relative) facilitate this bias. The variability evident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

A CONSTRUCTIVE APPRAISAL FOR FUTURE NI TRIALS

As previously discussed, too wide of an NI margin translates into excessive tolerability for the experimental treatment, whereas too narrow of an NI margin may be too tough of a hurdle, prohibiting some truly noninferior 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

FIGURE 2 Forest Plot of CIs for Absolute Differences and Relative Risks From 9 Stent Trials



Inverted triangle represents the pre-specified absolute and relative NI margins. ABSORB III = A Clinical Evaluation of Absorb BVS, the Everolimus Eluting Bioresorbable Vascular Scaffold in the Treatment of Subjects With de Novo Native Coronary Artery Lesions; ABSORB Japan = A Clinical Evaluation of AVJ-301 (Absorb BVS), the Everolimus Eluting Bioresorbable Vascular Scaffold in the Treatment of Subjects With de Novo Native Coronary Artery Lesions in Japanese Population; BIOSCIENCE = A Randomized Comparison of a Sirolimus-eluting Stent With Biodegradable Polymer Versus an Everolimus-eluting Stent With a Durable Polymer for Percutaneous Coronary Revascularization; CENTURY II = Clinical Evaluation of New Terumo Drug-Eluting Coronary Stent System in the Treatment of Patients with Coronary Artery Disease; COMPARE II = Comparison of the Everolimus Eluting (XIENCE-V, XIENCE-Prime or PROMUS Stent) With the Biolimus A9 Eluting NOBORI Stent in All-comers: a Randomized Open Label Study; DUTCH PEERS = DURable polymer-based sTent CHallenge of Promus ElemEnt versus ReSolute integrity in an all comers population; PLATINUM = A Prospective, Randomized, Multicenter Trial to Assess an Everolimus-Eluting Coronary Stent System (PROMUS Element) for the Treatment of up to Two De Novo Coronary Artery Lesions; RESOLUTE-AC = RESOLUTE-III All-comers Trial: A Randomized Comparison of a Zotarolimus-Eluting Stent With an Everolimus-Eluting Stent for Percutaneous Coronary Intervention; SORT OUT VI = Randomized Clinical Comparison of Biomatrix Flex and Resolute Integrity Coronary Stents in Non-selected Patients With Ischemic Heart Disease; other abbreviations as in [Figure 1](#).

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 among experts or agencies. Similar to how an NI margin that is too narrow makes 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 stent trials discussed earlier should raise concerns about the deliberate use of absolute margins and its consequences. Active research areas that repeatedly use NI designs, such as coronary stents, require a better consensus regarding the consistent choice of NI margins.

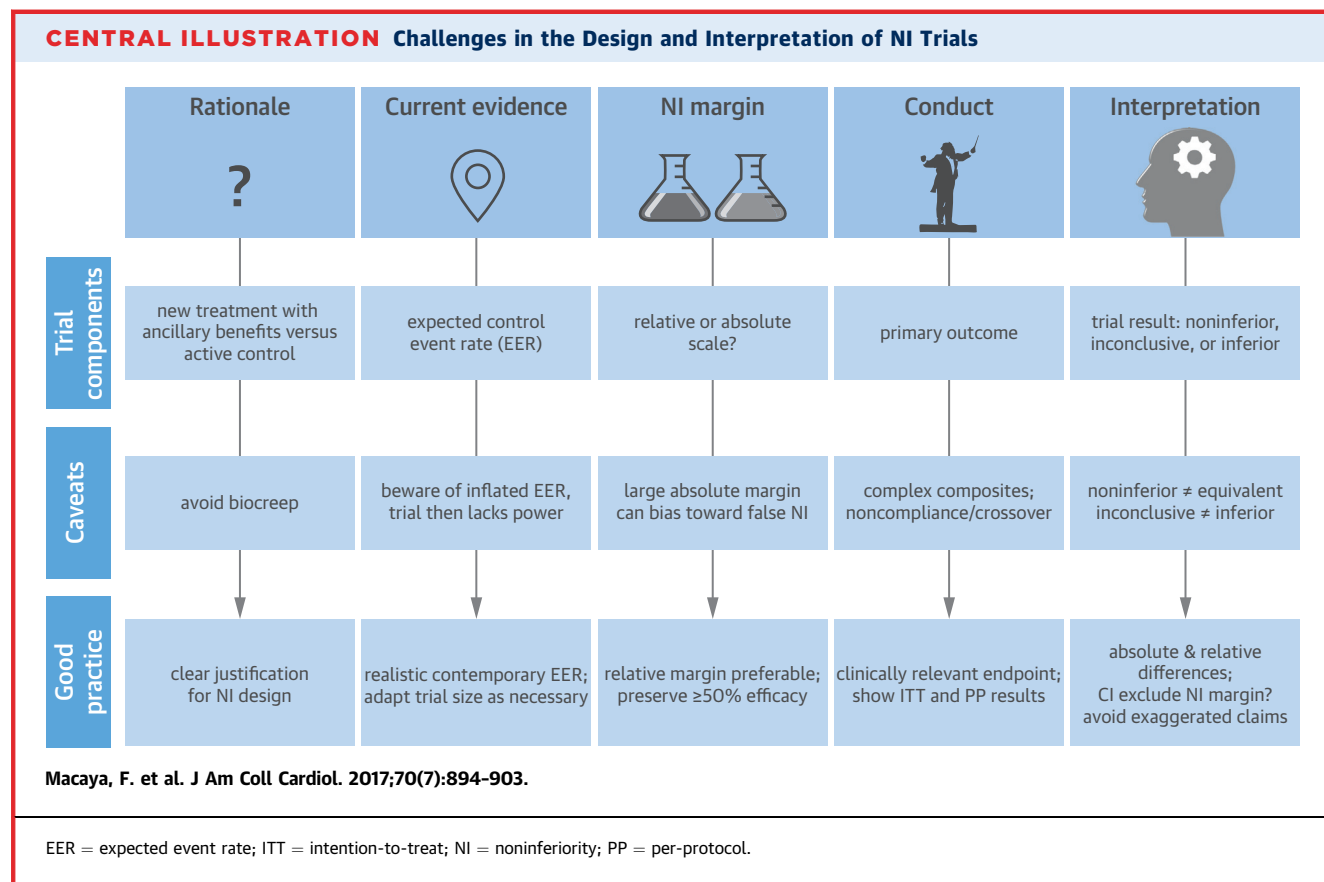
In contrast, NI trials evaluating drugs tend to use a relative rather than an absolute risk when choosing an NI margin. A controversy over the potential risk of myocardial infarction in patients with diabetes who were taking rosiglitazone (eventually not confirmed) led to the FDA and EMA mandating cardiovascular safety outcomes for all diabetic medications. Following this 2007 mandate, numerous NI 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](#)).

TABLE 3 ITT and PP or AT Analyses From the Stent Trials

Study Name	% NI Margin (RR)	1-Sided CI for ITT Analysis		1-Sided CI for PP/AT Analysis	
		95% CI	97.5% CI	95% CI	97.5% CI
RESOLUTE AC	3.5 (1.44)	1.8 (1.24)*	2.2 (1.30)	2.0 (1.29)	2.3 (1.35)
PLATINUM	3.5 (1.64)	1.9 (1.78)*	2.2 (1.95)	2.0 (1.88)	2.3 (2.06)
COMPARE II	4.0 (1.42)	1.8 (1.44)*	2.1 (1.52)	1.9 (1.52)	2.2 (1.62)
CENTURY II	5.5 (1.55)	1.7 (1.42)*	2.1 (1.55)	1.5 (1.39)	1.9 (1.52)
BIOSCIENCE	3.5 (1.44)	1.6 (1.28)*	2.0 (1.35)	1.2 (1.21)	1.6 (1.27)
DUTCH PEERS	3.6 (1.36)	2.7 (1.61)*	3.0 (1.71)	2.7 (1.63)	3.0 (1.74)
SORT OUT VI	2.5 (1.38)	1.6 (1.36)*	1.8 (1.43)	Not presented	Not presented
ABSORB III	4.5 (1.64)	3.6 (1.72)	4.0 (1.82)*	3.8 (1.75)	4.2 (1.85)
ABSORB Japan	8.6 (1.96)	3.8 (2.63)*	4.4 (3.11)	Not presented	Not presented

The pre-defined absolute NI margin used in the designs and their relative proportion to the expected event rate are presented. The 95% and 97.5% upper CI limits for the difference and for the RRs in parentheses were derived from ITT or PP/AT analyses. *Values used for testing noninferiority by each study.

AT = as-treated; CI = confidence interval; ITT = intention-to-treat; PP = per-protocol; other abbreviations as in [Table 2](#).



Thankfully, nowadays we observe relatively low primary event rates with most drug and device trials in cardiology. However, this provides a real challenge for NI 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 downward 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 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 an NI trial, the investigator should reflect on the need for an NI design and what the new treatment adds to our knowledge and patient benefit. Several late-generation stents developed over the past decade have, for the most part, been assessed in NI trials compared with Xience V (Abbott Vascular, Santa Clara, California). 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., bioresorbable 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 an NI trial do not rely entirely on a test for NI or an upper CI 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 considered to draw meaningful conclusions from each NI trial (**Central Illustration**).

Noninferiority trials can be particularly sensitive to bias due to deficiencies in their design and execution. As discussed earlier, 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, NI designs have an important role in cardiovascular therapeutics, and especially in evaluating new coronary stents. We hope our methodological perceptions, illustrated by recent examples, are of value in enhancing the statistical/scientific quality of future NI trials.

ACKNOWLEDGMENTS The authors are grateful to Alfonso de Hoyos, PhD, for his help creating the forest plots. Dr. Macaya thanks the Department of Preventive Medicine and Public Health, Facultad de Medicina—Universidad de Navarra, which has helped him and many other physicians gain a deeper knowledge and understanding of statistics.

ADDRESS FOR CORRESPONDENCE: Dr. Pablo Salinas, Hospital Clínico San Carlos, Calle del Profesor Martín Lagos, s/n. 28040, Madrid, Spain. E-mail: salinas.pablo@gmail.com.

REFERENCES

- Gopal AD, Desai NR, Tse T, Ross JS. Reporting of noninferiority trials in ClinicalTrials.gov and corresponding publications. *JAMA* 2015;313:1163–5.
- Vaduganathan M, Patel RB, Samman-Tahhan A, et al. Cardiovascular clinical trials with non-inferiority or equivalence designs from 2001 to 2012. *Int J Cardiol* 2016;214:16–8.
- 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.
- Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet* 2007;370:1875–7.
- Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med* 2006;145:62–9.
- Pocock SJ, Clayton TC, Stone GW. Challenging issues in clinical trial design: part 4 of a 4-part series on statistics for clinical trials. *J Am Coll Cardiol* 2015;66:2886–98.
- Collier TJ. Tools & techniques—statistics: non-inferiority trials. *EuroIntervention* 2014;10:526–7.
- Leon MB, Smith CR, Mack M, et al., for the PARTNER Trial Investigators. Transcatheter aortic-valve implantation for aortic stenosis in patients who cannot undergo surgery. *N Engl J Med* 2010;363:1597–607.
- Smith CR, Leon MB, Mack MJ, et al., for the PARTNER Trial Investigators. Transcatheter versus surgical aortic-valve replacement in high-risk patients. *N Engl J Med* 2011;364:2187–98.
- Leon MB, Smith CR, Mack MJ, et al., for the PARTNER 2 Investigators. Transcatheter or surgical aortic-valve replacement in intermediate-risk patients. *N Engl J Med* 2016;374:1609–20.
- Marso SP, Daniels GH, Brown-Frandsen K, et al., for the LEADER Steering Committee, LEADER Trial Investigators. Liraglutide and cardiovascular outcomes in type 2 diabetes. *N Engl J Med* 2016;375:311–22.
- Ellis SG, Kereiakes DJ, Metzger DC, et al., for the ABSORB III Investigators. Everolimus-eluting bioresorbable scaffolds for coronary artery disease. *N Engl J Med* 2015;373:1905–15.
- 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.
- Fifth Organization to Assess Strategies in Acute Ischemic Syndromes Investigators. Comparison of fondaparinux and enoxaparin in acute coronary syndromes. *N Engl J Med* 2006;354:1464–76.
- Green JB, Bethel MA, Armstrong PW, et al., for the TECOS Study Group. Effect of sitagliptin on cardiovascular outcomes in type 2 diabetes [Published correction appears in *N Engl J Med* 2015;373:586]. *N Engl J Med* 2015;373:232–42.
- Kuck KH, Fürnkranz A, Chun KRJ, et al., for the FIRE AND ICE Investigators. 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.
- Kimura T, Kozuma K, Tanabe K, et al., for the ABSORB Japan Investigators. 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.
- Raungaard B, Jensen LO, Tilsted HH, et al., for the Scandinavian Organization for Randomized Trials with Clinical Outcome (SORT OUT). 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;385:1527–35.
- 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;383:413–23.
- 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.
- Saito S, Valdes-Chavarrí M, Richardt G, et al., for the CENTURY II Investigators. A randomized, prospective, intercontinental evaluation of a bioresorbable polymer sirolimus-eluting coronary stent system: the CENTURY II (Clinical Evaluation of New Terumo Drug-Eluting Coronary Stent System in the Treatment of Patients with

- Coronary Artery Disease) 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; 381:651–60.
23. Stone GW, Teirstein PS, Meredith IT, et al., for the PLATINUM Trial Investigators. A prospective, randomized evaluation of a novel everolimus-eluting coronary stent: The PLATINUM (A Prospective, Randomized, Multicenter Trial to Assess an Everolimus-Eluting Coronary Stent System [PROMUS Element] for the Treatment of up to Two De Novo Coronary Artery Lesions). *Trial. 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; 363:136–46.
25. D'Agostino RB Sr., 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., for the U.S. CoreValve Clinical Investigators. 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., for the PROCEED II Trial Investigators. Ex-vivo perfusion of donor hearts for human heart transplantation (PROCEED II): a prospective, open-label, multicentre, randomised non-inferiority trial. *Lancet* 2015;385:2577–84.
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;308:2594–604.
29. Office of Biostatistics and the Office of New Drugs in the Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER), Food and Drug Administration. Non-inferiority clinical trials to establish effectiveness. Guidance for industry. November 2016. Available at: <https://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf>. Accessed June 13, 2017.
30. European Medicines Agency, Committee for Medicinal Products for Human Use (CHMP). Guideline on the choice of the non-inferiority margin. July 2005. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003636.pdf. Accessed June 13, 2017.
31. Le Henaff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA* 2006; 295:1147–51.
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;383:386–8.
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.

KEY WORDS clinical trial, clinical trial design, coronary stent, noninferiority margin, sample size