***Interprof. J. Health Sci.* 2019, *17(2)*: *114-136*.**
ISSN (online) 2672-9423 ISSN (print) 2672-9628

WHAT A PHARMACIST/PRACTITIONER SHOULD KNOW ABOUT EVALUATING NONINFERIORITY TRIALS

Quan Truong, Supakit Wongwiwatthananukit* and Erik Ferreira

Daniel K. Inouye College of Pharmacy, University of Hawaii at Hilo, Hilo, Hawaii, USA

*Corresponding author : E-mail : supakit@hawaii.edu
Received 5 July 2019; Accepted 20 August 2019

Abstract: Noninferiority trials attempt to prove noninferiority of test therapy over existing standard treatment and are a prominent component of modern drug testing and approval. It is essential that pharmacists and practitioners of evidence-based medicine can critically evaluate these studies for strengths, weaknesses, and applicability to patient care. This review aims to present a question-based (questionnaire) approach of how to evaluate the minutiae of noninferiority trials, adapted from CONSORT (Consolidated Standards of Reporting Trials) Guidelines and its extensions for evaluating noninferiority trial and other regulatory guidance documents. An illustrative example of using such a questionnaire is provided to help practitioners evaluate a published noninferiority trial prior to applying the results to patient care.

Keywords: Noninferiority trial, evidence-based medicine, pharmacist, CONSORT

INTRODUCTION

Conventionally, clinical trials attempt to demonstrate the superiority of an intervention over the control, most often placebo, hence the term superiority trials. Meanwhile, a noninferiority (NI) trial is a clinical study in which researchers seek to compare an experimental intervention with an active control, often the existing standard treatment, to show that the new intervention is "not inferior", or not clinically worse than the active control (Head *et al*, 2012; Piaggio *et al*, 2012; Schumi and Wittes, 2011; U.S. FDA, 2016). When comparing the experimental treatment to the control in NI trials, researchers choose an acceptably small margin of difference called the NI margin. If the study outcome falls within this margin, the difference between experimental treatment and the control would be small enough to conclude that the experimental data is considered "noninferior" to the control. In other words, the new treatment is also effective, given its "not-less-effective" status relative to the active control. Noninferiority studies are useful for several reasons: Firstly, it may not be possible for researchers to demonstrate superiority of the new treatment over existing standard treatment. While being not superior to the control, new therapy might offer ancillary benefits that would justify its use, for instance, better safety profiles, improved convenience, good tolerability, and less costly. In addition, NI trials are used in situations in which a placebo trial would be considered problematical, even unethical because there is an existing effective medical treatment, for example, a trial to evaluate new antibiotics for curing infection (Bryant and McQueen, 2018).

The use of NI clinical trials has been on the rise, with just one NI trial reported in 1998 versus 583 articles incorporating NI-designed studies in 2009, a dramatic increase in just over ten-year period (Suda *et al*, 2011). Moreover, among 43 approved New Drug Applications (NDAs) between 2002 and 2009, two-thirds contained evidence obtained from NI studies. Nevertheless, NI clinical trials are not without challenges, especially regarding the design and method of the studies. For a NI trial to be valid methodologically, among many other requirements, the researchers need to explicitly state the NI margin and justify the selection of such a margin based on previous trials or by other validated means (Henanff *et al*, 2006; Mauri and D'Agostino, 2017; U.S. FDA, 2016). Prior to 2010, before the draft of U.S. FDA guidance on NI studies, many NI trials failed to report the margin of NI (Henanff *et al*, 2006; Piaggio *et al*, 2012); from 2010 to 2015, among 164 NI trials published in major medical journals that reported NI margins, more than half did not provide justification for such margins (Rehal *et al*, 2016). Methodological flaws in design and inconsistency in reports of NI studies render NI conclusions biased and misleading, directly affecting applicability of such results to patient care. Thus, the responsibility lies upon pharmacists and practitioners to discern the usefulness of NI studies in order to apply to patient care in accordance with evidence-based medicine principles. The purpose of this review is to provide pharmacists, health care practitioners, and students with a comprehensive, questionnaire approach to evaluate a NI trial and apply the results to patient care.

Table 1 features a question-based approach to evaluate noninferiority (NI) trials, adapted from Extension of the CONSORT 2010 Statement, U.S. FDA 2016 Guidance for Industry, EMA 2000 and 2006 statement, EPC Working Group AHRQ 2012, and from other resources. This approach will allow readers to evaluate NI trials by breaking the study into sections, scrutinizing each section with a series of questions. Readers should also keep in mind that NI clinical trials are, in essence, clinical studies and thus should be treated as such, with the same considerations regarding accurate and sound research design, method of randomization, description of follow ups, appropriate use of statistical methods, conflict of interest disclosures, etc. This review aims to highlight the main features of NI studies to which pharmacists, practitioners, and students should pay more attention in deciding the applicability of such evidence. At the end of the discussion, an illustrative example of using the questionnaire to evaluate the MAGELLAN study will be provided. Throughout the review, the term experimental intervention will indicate the new therapy or drug being evaluated, and "active control" will indicate the reference or standard treatment or active comparator to which the experimental intervention is compared.

Table 1. Question-based approach for evaluating noninferiority (NI) trials

Section/topic	*Questions to evaluate*
Title, Abstract, and Introduction	
Title and Abstract	Q. Are the title and abstract appropriate to identify the study as a NI trial? Could it be discerned which experimental intervention is being compared against which active control?
Introduction	Q. Was the rationale for using NI design provided? Was the NI hypothesis stated?
Methods	
Participants	Q. Were the participants in the NI trial similar to those in trial(s) that established efficacy of the active control?
Interventions	Q. Was the active control appropriately chosen? Was the active control in the NI trial identical or similar to that of historical trial(s) that established its efficacy? Q. Were assigned treatments carried out appropriately?
Outcomes	Q. Were the NI outcomes stated? Were primary and secondary outcomes based around NI or superiority hypotheses? Were the primary outcomes of NI identical or similar to that of historical trial(s) that established efficacy for the active control?
Sample size and NI margin	Q. Was the sample size calculated using NI criteria? Q. Was the NI margin specified *a priori* with an appropriate rationale? Q. Was the NI margin reported in an absolute or a relative scale?
Statistical Analysis	Q. Was an intention-to-treat (ITT) or per-protocol (PP) analysis used? Was missing data appropriately handled? Q. Was a one-sided or two-sided confidence interval approach used?
Results and Application to Patient Care	
Participant characteristics	Q. Was prognostic balance between experimental and control groups maintained before, during, and at completion of the trial?
Outcomes and Estimation	Q. Were the study results of NI outcomes presented in relation to confidence intervals and NI margin? Q. Were sensitivity analyses done to test for robustness of results?
Discussion	Q. Were the results interpreted in relation to the NI hypothesis? Was conclusion of NI warranted? Q. Was the effect of active control preserved? Q. If a superiority conclusion was drawn for the outcomes for which NI was hypothesized, did the authors provide justification for switching methods?
Applicability to patient care	Q. Are my patients similar to the study participants? Q. Are the benefits of using noninferior intervention worth the trade-off in effectiveness against not using standard treatment? Is the loss of efficacy acceptable to my patients?

(Adapted from EMA 2001; EMA 2006; Kaul *et al*, 2006; Piaggio et al, 2012; U.S. FDA, 2016)

Title, Abstract, and Introduction

Q. Are the title and abstract appropriate to identify the study as a NI trial? Could it be discerned which experimental intervention is being compared against which active control?
Q. Was the rationale for using a NI design provided? Was the NI hypothesis stated?

A NI trial should, at the minimum, indicate the study is a NI design if the primary outcome of the study is a NI outcome (Piaggio *et al*, 2012). The abstract should clearly state which intervention is being compared against which active control. This allows practitioners to determine whether this study matches what they are looking for. The introduction and background of the study should explain, in detail, the rationale of choosing the NI design over a superiority design. In other words, the researchers should describe the standard treatment and the drawbacks of using it, which in turn should be used as the active control in the study. In addition, the background should describe the experimental intervention in terms of ancillary benefits it demonstrates over the active control.

Noninferiority studies, like any other studies, attempt to formulate and answer the research question about the experimental intervention, and thus the NI hypothesis should be explicitly stated, preferably with the NI margin included (Piaggio *et al*, 2012). Table 2 compares the hypothesis testing of superiority studies and NI studies. In short, the researchers try to prove the alternative hypothesis of NI studies, which means that the difference between the experimental and the control is less than the pre-specified margin of NI, i.e., the experimental is not worse than the standard treatment by an amount not larger than the NI margin. The NI margin will be discussed further in Method question. Essentially, practitioners should examine the background of the study carefully to capture the basis of the study and understand the NI hypothesis before progressing further.

Table 2. Comparison of hypothesis testing between superiority and noninferiority trials

Superiority trial	Noninferiority trial
Ho: $T \leq C$ or $T - C \leq 0$	Ho: $C - T \geq M$
Ha: $T > C$ or $T - C > 0$	Ha: $C - T < M$

Ho = null hypothesis; Ha = alternative hypothesis; T = test therapy; C = control; M = margin of noninferiority
(Adapted from U.S. FDA 2016)

For superiority or placebo-controlled studies, the null hypothesis states that the difference between test treatment and control (or placebo) is less than zero, while the alternative hypothesis states that the difference is more than zero. For NI studies, the null hypothesis states that the difference between the active control and the experimental test is more than the NI margin, while the alternative hypothesis states that the difference is less than the NI margin, or the experimental test is inferior to the active control by less than the NI margin.

Methods – Participants

Q. Were the participants in the NI trial similar to those in trial(s) that established efficacy of the active control?

Participants in the NI trial of interest should have characteristics similar to those of previous trials which examined the efficacy of the active control over placebo or other active comparators, for example, similar inclusion and exclusion criteria, similar baseline co-morbidities, similar use of pharmacological agents, etc. Such criterion, called constancy assumption, is to ensure that the active control arm preserved its efficacy in the new study, a concept known as assay sensitivity, a unique yet important feature of NI studies (see Figure 1) (Piaggio *et al*, 2012; U.S. FDA, 2016). In other words, when designing a NI study, researchers must assume it is essential that the active control retains its comparative effectiveness in the new context of the trial; thus, keeping study participants and conditions similar to historical trials. Readers should look closely at the description of participants and make note of the authors' efforts to maintain the constancy assumption in the study, which is not a straightforward process, especially if researchers did not describe the details. In that case, readers are encouraged to locate the reference trials and explore the similarities of the two participant pools. Discordance of current study participants compared with historical trial participants deserves a detailed explanation from the authors, or else, the study results may be biased toward the NI conclusion if the efficacy of the active control is not well maintained.

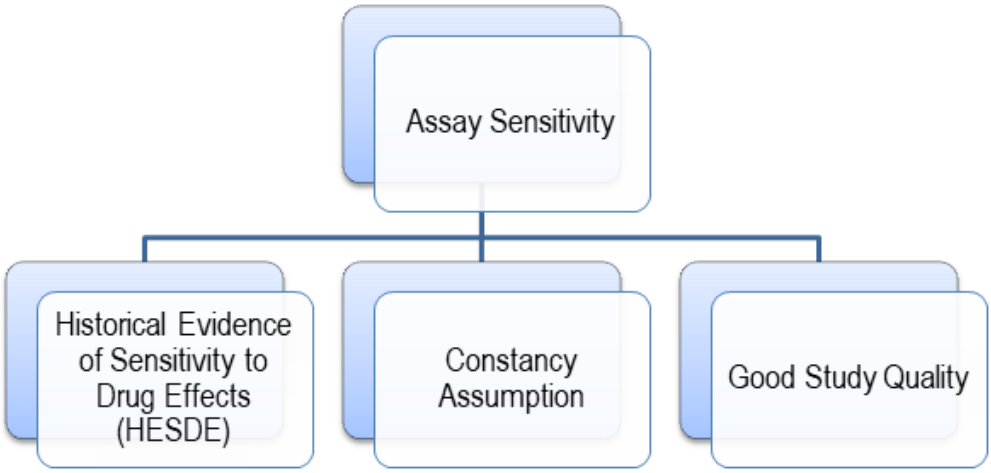

Figure 1. Assay Sensitivity of a Noninferiority Trial (Adapted from U.S. FDA, 2016)

Assay Sensitivity is the ability of the study to detect a treatment difference between the active control and ineffective placebo as if the NI studies include a placebo arm, aka preserved its efficacy over placebo (active control to be superior to placebo) in the current NI trial. Assay Sensitivity is determined by the following three factors: Historical Evidence of Sensitivity to

Drug Effects (HESDE), Constancy Assumption, and Good Study Quality. HESDE is the consistent treatment effect of the active control over placebo in trials that established the active control efficacy. Constancy Assumption is the conclusion that HESDE is reached only when the NI studies have adequately similar characteristics to those of historical trials. Good Quality Study ensures the internal validity and to guard against bias to the alternative hypothesis of noninferiority.

Methods – Interventions

Q. Was the active control appropriately chosen? Was the active control in the NI trial identical or similar to that of historical trial(s) that established its efficacy?
Q. Were assigned treatments carried out appropriately?

Active control in any NI studies should be the acceptable standard treatment at the time of the trial, since selection of a less effective active control would make the experimental intervention perform look less inferior. (Piaggio *et al*, 2012; U.S. FDA, 2016). Readers should be aware of the biocreep concept, which is a problem of increasing use of NI clinical trial design (Vermeulen, 2011). For example, drug A has been confirmed NI as the gold standard with drug B in a NI trial, and due to drug A being used more, it is now used as the active control to test for NI of a new drug C, the experimental intervention. In that case, even if drug C is shown to be NI to drug A, the experimental intervention may actually be truly inferior to the old standard, drug B, and, even worse, drug C might not perform better than placebo. Therefore, selection of active control should be correctly in the first place, or else, the NI trial loses its meaning.

On the same premise of constancy assumption (Figure 1), the active control in the current NI trial should be identical or sufficiently similar to that of previous trials that established its efficacy, in terms of dosing, frequency, settings of administration, etc. If the active control was underdose, the study results would be biased towards the NI hypothesis. In contrast, if the active control dose was increased, it would raise a tolerability problem for the control and falsely acclaim the experimental intervention's ancillary benefits (Piaggio *et al*, 2012). On the other hand, assigned treatments, both the active control and experimental intervention, should be carried out diligently according to current administration standards to ensure the NI study's internal validity. While poor internal validity of superiority trials may bias the results toward the null hypothesis, poor internal validity of NI trials may actually bias the results toward the alternative hypothesis of noninferiority. It is the responsibility of pharmacists and practitioners to thoroughly inspect NI trials for its internal validity, particularly with study treatments and interventions.

Methods – Outcomes

Q. Were the NI outcomes stated? Were primary and secondary outcomes based on NI or superiority hypotheses? Were the primary outcomes of NI identical or similar to that of historical trial(s) that established efficacy for the active control?

Obviously, the primary outcome of any NI studies should be an NI outcome, which would be used to calculate sample size (Piaggio *et al*, 2012). Secondary outcomes of NI studies may be

NI outcomes or superiority outcomes, but these outcomes should be specified a priori to avoid opportunistic bias. Once again, readers are called to assess the similarity of primary outcomes of NI to that of historical trials under the constancy assumption basis, such as similarity in types of outcome, the measuring device, measurement frequency, and timing of measurements. Because differences in outcome aspects compared with previous trials may falsely claim NI status of the experimental intervention, justification of such difference outcome measurements should be provided in the method section (Mauri and D'Agostino, 2017). Difference in composite outcomes between historical trials and a current NI trial, for example, must be reconciled in detail, especially if there is incorporation of any outcome with little clinical benefit.

Methods - Sample size and NI margin

Q. Was the sample size calculated using NI criteria (i.e. the noninferiority margin)?
Q. Was the NI margin specified *a priori* with a rationale?

Sample size of NI studies should be calculated using the primary NI outcome. Usually a larger sample size is required for NI studies compared to superiority trials (Kaji and Lewis, 2015). Sample size requirement is directly influenced by the choice of the NI margin, which is undoubtedly the most important information in a NI trial. A large margin of NI would risk a wrong conclusion of NI for a truly inferior experimental intervention (i.e., inflated type I error), while a very small margin of NI would require a very large sample size to adequately detect such a small difference (i.e., increased type II error), thus rendering the study impractical (EMA, 2006; Kaul and Diamond, 2006; Piaggio *et al*, 2012; U.S. FDA, 2016). Therefore, the choice of NI margin is critical. Unfortunately, there has not been a universal protocol for calculating and determining the NI margin, and even guidance from regulatory bodies are not consistent in their explanations. Thus, readers need to be vigilant in distinguishing between an acceptable NI margin and an unacceptable one, even more so when application to patient care is considered.

Recall that the NI margin is the largest amount of statistically and clinically acceptable extent that the experimental intervention is allowed to be less effective, while still having a treatment effect, compared to the active control (Kaul and Diamond, 2006; Piaggio *et al*, 2012; U.S. FDA, 2016). For instance, given a cardiovascular outcome such as risk of heart attack, a NI margin of 1.3 on the relative scale means that the experimental intervention is allowed to be 30% less effective than the active control to be considered noninferior. The question then becomes which degree of less effectiveness is acceptable, or in other words, how large could the NI margin be? The concept of historical evidence of sensitivity to drug effects (HESDE) allows researchers to assume that the active control has its efficacy in the current trial and, in turn, allows the NI margin to be derived from evidence of these historical trials. There are three common methods to determine the NI margin: the point estimate, fixed margin, and synthesis method (U.S. FDA, 2016; Althunian *et al*, 2017). In the point estimate model, evidence of efficacy of the active control was pooled from different historical placebo-controlled trials using random-effect meta-analysis to produce a single 95% confidence interval (CI), and the lower bound of 95% CI is the conservative maximum treatment effect of the active control assumed to be present in the NI trial. It is termed *dmax*. Using *dmax* as the NI margin would risk losing the entire efficacy of the active control and make the experimental intervention worthless. Therefore, to preserve the active control efficacy and still test for effectiveness of the experimental intervention, a fraction of efficacy preserved, *f*, is imposed. Choosing *f* is a matter of clinical judgement, and thus it could influence applicability of the noninferior treatment to patients. Conventionally, it has been

suggested that *f* is about 50% for cardiovascular, oncology, and thrombotic trials or trials that evaluate mortality and have a small difference of event rate, while antibiotic trials, which have larger event rate of curing effects, could have a larger *f* value of about 85-90% on a risk difference scale (Kaul and Diamond, 2006; U.S. FDA, 2016). The NI margin then could be calculated from *dmax* and *f* using formula $M = dmax*(1-f)$ for a risk difference scale and $M = dmax^{(1-f)}$ for a relative scale. A larger *f* would produce a more stringent NI margin.

Similar to the point estimate method, the fixed margin method also utilizes meta-analysis of historical evidence to formulate the NI margin (U.S. FDA, 2016). However, instead of using the lower bound of 95% CI as the *dmax*, the fixed margin method applies discounting to the maximum treatment effect to balance the effect of variances and uncertainties of historical studies (imperfect constancy assumption), hence a smaller *dmax* is used to compare with the point estimate model. Both the point estimate model and the fixed margin method provide a calculation for the NI margin *a priori* using historical evidence and hence is a more well-known method for NI margin selection. On the other hand, the synthesis method combines both historical evidence and treatment effects from the current NI trial to formulate altogether one 95% CI with the NI margin. Thus, the NI margin is not specified in advance; however, the researchers should indicate what fraction of efficacy preserved, *f*, is to be used to determine the NI margin (Table 3).

There is no perfect method to choose the NI margin, and researchers are free to use other methods to determine the NI margin. However, researchers should justify their choice in designing the study and dutifully report their analyses of historical evidence to come up with their method of NI margin selection based on statistical and clinical reasoning (Piaggio *et al*, 2012; Schumi and Wittes, 2011; U.S. FDA 2016;). With the exception of the synthesis method, the NI margin should be pre-determined and clearly stated, and the protocol for obtaining the margin should be reported with the trial or separately, in an appendix or protocol report, regardless of method used. While pharmacists and practitioners are not required to calculate the NI margin when assessing the NI trial, it is their responsibility to verify the validity of the method to obtain the NI margin with given information. Unsurprisingly, the choice of the NI margin is the most critical factor of a NI study design, as it plays a big role in determining the applicability of the experimental intervention to patient care.

Table 3. Comparison between Point Estimate Method, Fixed Margin Method, and Synthesis Method for NI Margin estimation

	Point Estimate Method	**Fixed Margin Method**	**Synthesis Method**
Maximum conservative treatment effect, *dmax*	Using historical evidence from previous trials to estimate *dmax*	Using historical evidence from previous trials with discounting on *dmax* for variance effect	Using both historical evidence from previous trials and data from current NI trial
Fraction of efficacy preserved, *f*	Clinical judgement	Clinical judgement	Clinical judgement Need to specify f in advance
Calculation of NI margin, *M*	M = *dmax**(1-*f*) (absolute) OR M = $dmax^{(1-f)}$ (relative)	M = *dmax**(1-*f*) (absolute) OR M = $dmax^{(1-f)}$ (relative)	Not applicable NI margin not calculated in advance

(Adapted from U.S. FDA, 2016 and Althunian *et al*, 2017)

Q. Were NI outcomes and NI margin reported in absolute or relative scale?

Choice of metric for the primary outcome should also be stated, either absolute risk difference or relative scale effect, since the interpretation of the NI conclusion may differ between the two choices of metric (U.S. FDA, 2016; Palmas, 2017). As mentioned, the NI margin should reflect the largest acceptable loss of efficacy for the active control in past trials; thus, reporting in relative scale guards against any variability between studies that might affect interpretation of the NI results, which makes relative risk, odd ratio, and hazard ratio a more favorable choice in NI trials. On the other hand, absolute risk difference allows for evaluation of both treatment arms in the current setting and is more useful in subsequent superiority testing, if any. Therefore, it is suggested that authors report and interpret NI margin using both absolute risk difference and relative scale. Conclusion of NI is strengthened if results of both absolute and relative scales agree, but in the case that the two scales do not agree, any NI conclusion must be interpreted carefully with emphasis on the relative scale.

Methods - Statistical Analysis

Q. Was a intention-to-treat (ITT) or per-protocol (PP) analysis used? Was missing data appropriately handled?
Q. Was a one-sided or two-sided confidence interval approach used?

The ITT analysis approach, which accounts for all patients who have been randomized to study groups regardless of dropout rate, loss to follow up, and accidental crossover, is recommended in superiority trials as it is the conservative determination of treatment effect, in which a poorly conducted superiority trial will bias the result toward the null hypothesis of non-significant difference (Mulla *et al*, 2012; Piaggio *et al*, 2012; U.S. FDA, 2016). In contrast, ITT analysis may bias the result towards the alternative hypothesis in a NI trial, by making the experimental intervention look less inferior when it is truly inferior to the active control. PP analysis which only takes into account participants who finished the study is thus considered the appropriate approach in NI trials, as it can guard against a false conclusion of noninferiority by ensuring the integrity of the active control arm. However, PP analysis may create problem for the overall validity of the trial by discarding the effect of randomization if there were too many dropouts (Mauri and D'Agostino, 2017). Therefore, it is recommended that researchers analyze data using both ITT and PP approaches; if results from both analyses do conclude NI, then the study result is more credible. An important aspect of an ITT analysis is in the handling of missing data. Therefore, the authors should report in details of what was done with missing data in the protocol section. Readers should be vigilant to any discordance between the results of ITT and PP analyses in a NI trial. Any discrepancy needs thorough justification from the authors.

Beside appropriate statistical tests used for analyzing primary and secondary outcomes, interpretation of NI trials is usually done with a one-sided test approach with 0.025 significance level for the primary NI outcomes, i.e., the one-end limit of a two-sided 95% CI (Piaggio *et al*, 2012; U.S. FDA 2016; Rehal *et al* 2016). Nevertheless, many NI studies indeed use one-sided significance level at 0.05 for the NI outcome, which corresponds to an upper bound (or lower bound) level of two-sided 90% CI, which brings into question the appropriateness of the statistical parameters of the study. Generally, this situation is appropriate only when there is a homogenous historical evidence of the active control's efficacy, with the authors providing detailed explanation of their choice of significance level. Meanwhile, a two-sided test approach

with alpha set at 0.05 is typically used after the NI hypothesis is confirmed to test for superiority, where 95% CI lies above zero (for risk difference) or 1 (for relative scale) will ensure superiority of the experimental intervention over the active control, if the protocol specifically allows this sequential testing.

Results - Participants' Characteristics

Q. Was prognostic balance between experimental and control groups maintained before, during, and at completion of the trial?

Similar to other types of clinical trials, baseline characteristics of intervention and control arms of NI studies should be balanced and maintained throughout the study to protect the study's internal validity (Mulla *et al*, 2012). Given its hypothesis testing, NI clinical trials are particularly susceptible to misleading conclusions when inherent bias is not well controlled, and the imbalance of baseline characteristics is problematic. This may diminish the active control's efficacy and thus make the experimental intervention appear to perform better. Therefore, pharmacists and practitioners should pay attention to reports of study flow, study participants, dropout rates, and how those could influence the balance between the two study groups when analyzing a NI trial. Any substantial difference that may influence the result deserves a detailed explanation and, if necessary, sensitivity analyses to rule out possible false NI conclusion.

Results - Outcomes and Estimation

Q. Were the study results of NI outcomes presented in relation to confidence intervals and NI margin?
Q. Were sensitivity analyses done to test for robustness of results?

Presentation of results with regard to NI outcomes is best done with figures depicting the 95% CI and the NI margin (Piaggio *et al*, 2012). If the upper bound of the 95% CI lies well below the NI margin (for undesirable outcomes such as mortality or morbidity), the experimental intervention is noninferior to the active control by the NI margin, and the reverse is true for desirable outcomes. When the 95% CI crosses the NI margin, NI could not be demonstrated for experimental intervention because the loss of efficacy compared to active control is excessive (Figure 2). The p-value for NI comparison is also used to determine statistical significance, but it is a less preferable method. Report of different analyses should also be presented, preferably in the format of a figure, for example to show the difference between ITT and PP analysis. Because NI trials are prone to bias toward NI conclusion, sensitivity analyses should be done to examine the robustness of the results, looking at the effect of missing data, dropout rates, any difference in baseline characteristics, etc. (Rehal *et al*, 2016). Readers are encouraged to note any significant results from sensitivity analyses and verify the authors' rationale of the NI conclusion, if any, in lieu of these analyses. Other results regarding superiority or safety outcomes should be reported in the format of a table for ease of interpretation.

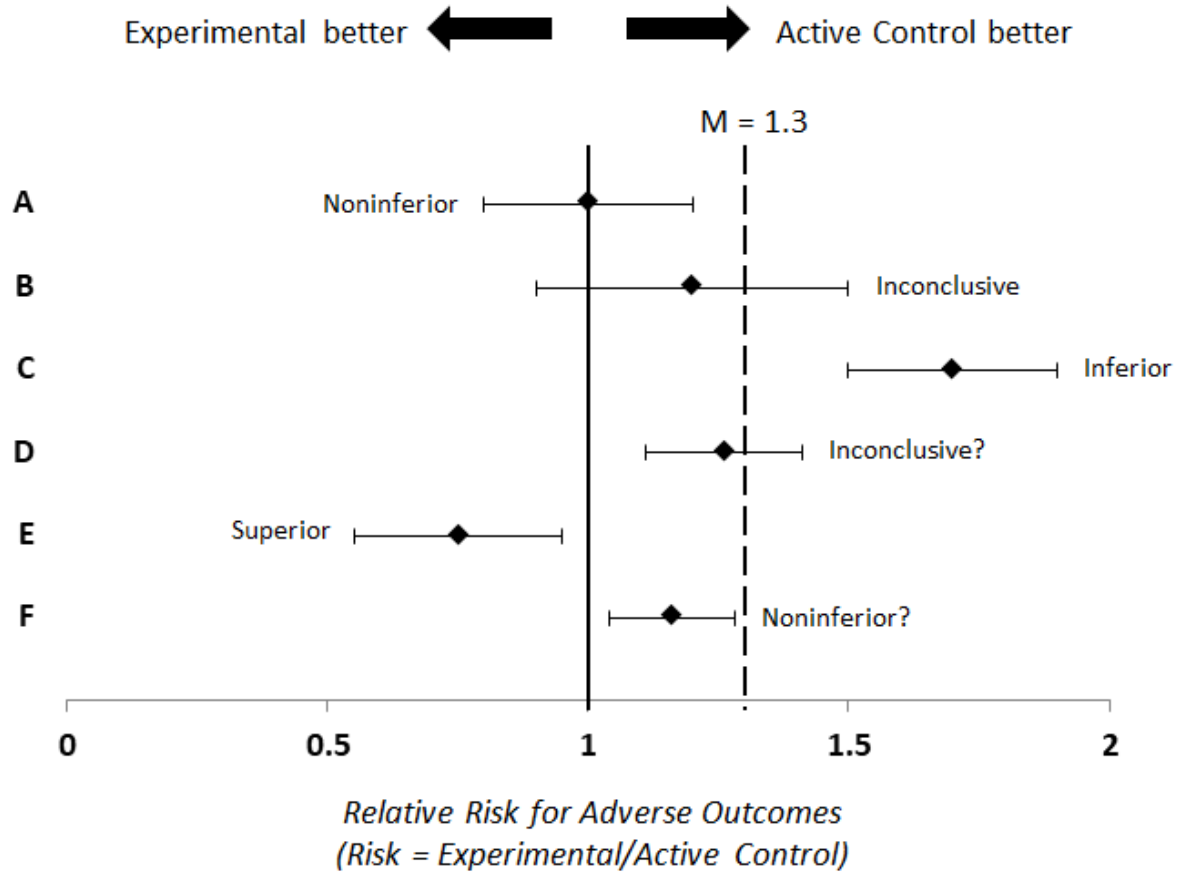

Figure 2. Interpretation of Possible Results of an NI Study Using 95% CI in Relation to NI Margin, *M*

A. Upper bound of 95% CI is 1.2 and below NI margin, *M*; NI is demonstrated.
B. Upper bound of 95% CI is 1.5 and above *M*; thus, NI is not demonstrated.
C. The entire 95% CI is well above *M*; thus the experimental is indeed inferior to the active control.
D. The upper bound of 95% CI is 1.41 and above *M*, so NI is not demonstrated. However, the experimental performs significantly worse than the active control, evidenced by the lower bound of 95% CI above being above 1.
E. The entire 95% CI is below *M* and below 1; both NI and superiority is demonstrated
F. While the upper bound of 95% is below *M* and NI is indicated, but lower bound 95% CI is above 1, suggesting that the experimental intervention is significantly worse and truly inferior to the active control.
(Adapted from Head *et al*, 2012; Piaggio *et al*, 2012; Schumi and Wittes, 2011; U.S. FDA, 2016)

Results - Discussion

Q. Were the results interpreted in relation to the NI hypothesis? Was conclusion of NI warranted?
Q. Was the effect of active control preserved?

Interpretation of results should be based on the primary outcomes of NI, and readers are encouraged to use a conservative approach in interpreting the result of any NI trials despite the

authors' conclusion (Mulla *et al*, 2012). Recall that NI study results should be interpreted with regard to absolute and relative scales, and disagreement between the results of two scales signify that NI is not demonstrated in the trial. Similarly, disagreement between ITT and PP analyses may jeopardize the NI conclusion of experimental intervention. Results of sensitivity analyses are also useful in assessing the confidence of NI conclusion. Nevertheless, above these all, the preserved efficacy of active control represents the most critical information to warrant the NI conclusion (Mulla *et al*, 2012; U.S. FDA, 2016; Mauri and D'Agostino, 2017). In other words, the active control should achieve adequate efficacy as demonstrated in previous trials, ideally the conservative maximum treatment effect (*dmax*). Assessment of this is not possible in many cases, as a placebo is not included in the study. As discussed, a poorly conducted active control arm might result in a sub-optimal effect of the active control and thus make it less effective. Moreover, difference in participants' characteristics, poor randomization, high attrition rate, etc., might have an effect of the active control performance. If, for any reason, the active control efficacy is vastly smaller than its historical evidence, a claim of NI is a false conclusion, as doing so will lead practitioners inadvertently to use a truly inferior therapy inadvertently.

Q. If a superiority conclusion was drawn for the outcomes for which NI was hypothesized, did the authors provide justification for switching?

Superiority for the experimental intervention could also be claimed in a NI trial, provided that the protocol justifies the switch *a priori* (EMA, 2001). Recall that secondary outcomes of NI trials are not restricted to NI outcomes, and thus when primary NI outcomes are confirmed, secondary outcomes of superiority might be explored according to a hierarchical testing premise. Nevertheless, there are requirements for switching the objectives from NI to superiority, and the protocol and statistical design should be specified in advance, including requirements of proper study design with regards to power and treatment effects. ITT can be the main analysis for superiority testing, and the p-value for superiority set at 0.05 for a two-sided 95% CI interval, among other criteria applied for conventional superiority clinical trials. Generally, if the 95% CI lies does not cross zero (for absolute risk difference) or does not cross 1 (for relative scale), superiority of the experimental intervention is demonstrated over the control, so long as other conditions are achieved and maintained. An example of objective switching from NI to superiority testing is the use of a NI trial to demonstrate superiority for cardiovascular (CV) safety of new antidiabetic medications (U.S. FDA, 2008; Campbell-Scherer, 2017). After demonstrating NI for CV safety as an add-on therapy to standard treatment, the new medications can be tested for superiority of CV safety if there is a protocol for such sequential testing. While switching objectives from NI to superiority is commonly done, the reverse situation with superiority trials is usually not acceptable since the sensitivity of a superiority trial is not designed to detect such a small difference in the NI margin, if any. Under rare circumstances, this reverse switching is allowed with strict requirements (Table 4).

Table 4. Points to Consider When Switching Objectives trials between NI and Superiority

Requirements when Switching Trials' Objectives from NI to Superiority	Requirement when Switching Trials' Objectives from Superiority to NI
● Properly designed trials which were carried out with strict requirements of NI studies ● *A priori* determined superiority hypothesis ● P-value for superiority with two-sided 95% CI presented ● Emphasized ITT analysis	● NI margin determined in advance with respect to active control; in case the NI margin is not determined *a priori*, detailed justification must be provided ● Similar conclusions from ITT and PP analyses with corresponding 95% CI and p-values ● Properly designed trials which were carried out with the strict requirements of NI studies ● Adequate power calculation obtained by means of NI margin or similar assumptions. ● Evidence of preserved treatment effect of active control

(Adapted from EMA, 2001)

Results - Applicability to Patient Care

Q. Are my patients similar to the study participants?
Q. Are the benefits of using noninferior intervention worth the loss of efficacy trade-off from not using standard treatment? Is the loss of efficacy acceptable for my patients?

After accepting the conclusion that the experimental intervention is noninferior to the active control, pharmacists and clinicians are faced with another challenge: how appropriately apply this to patient care (Mulla *et al*, 2012). Similar to superiority trials, applicability of NI trials to treat patients would be mostly appropriate if patients match the description of participants, namely in age, comorbidities, duration of illness, safety issues, etc. The better the match, the more the benefit from the intervention. Readers of NI trials are also called to assess the risk-benefit ratio of using the noninferior intervention compared to the standard treatment. In other words, the benefits of using the noninferior therapy should, at the minimum, justify the loss of effectiveness of not using the standard treatment. Benefits of cost effectiveness, convenience of dosing, improved safety profile, and better quality of life should all be considered relative to the loss of efficacy.

In consideration of a patient's best interest, pharmacists and practitioners should independently question the loss of efficacy, using the upper or lower bound of 95% CI for comparison with the NI margin. Given a patient's condition, clinicians should determine at what level of loss of efficacy would the benefits of using noninferior therapy outweigh the increased risk of not using the standard treatment, and such a decision can vary from patient to patient. For example, the results of an NI study showed that the 95% CI is from 0.8 - 1.3, well within the NI margin of 1.4 on the relative scale for composite CV outcome. A decision to use the noninferior intervention, thus, would risk acceptance of 30% efficacy loss compared to the standard

treatment, which could be fairly unacceptable for some patients. Therefore, after careful consideration of evidence of effectiveness, patients' profiles, and clinicians' expertise, the decision of using the noninferior intervention should almost always be a shared decision making by both clinicians and patients.

Glossary Box of Common Concepts in Noninferiority Clinical Studies

Noninferiority (NI) trials or studies: active controlled trials intended not to demonstrate superiority of experimental intervention but to prove that it is not worse than the active control by a margin.

NI Margin: an acceptable largest amount of difference the experimental intervention allowed to be worse than the active control and still be considered noninferior; the maximum clinical loss of efficacy of the active control in the current trial; determined by statistical and clinical reasoning using various methods.

Experimental intervention: test drug or new therapy.

Active control: standard treatment or therapy.

Biocreep: using a noninferior drug or therapy as the active control instead of standard treatment.

Assay Sensitivity: ability of the NI trial to detect a difference between the active control and placebo, if the study included a placebo arm; the ability to preserve its efficacy over placebo.

Historical Evidence of Sensitivity to Drug Effects (HESDE): consistent treatment effect of the active control over placebo in trials that established the active control efficacy.

Constancy Assumption: conclusion that HESDE is reached only when the NI studies have adequately similar characteristics to those of historical trials; depending on similarity of participants' characteristics, concurrent therapies, study outcomes, active control dose and duration, entry criteria, etc.

Maximum conservative treatment effect, *dmax:* the largest treatment effect of active control based on conservative analysis of historical evidence; a parameter important in determining the NI margin.

Discounting: reasonable reduction of *dmax* due to variance effect of different study; used in Fixed Margin method

Fraction of preserved efficacy, *f:* amount of efficacy of active control clinically preserved in the NI trial; determined by clinical judgement; the larger the fraction, the more stringent the NI margin.

Bias Toward the Alternative Hypothesis: biases in study design and methodology that make the active control perform poorly and thus make experimental intervention look falsely noninferior.

ILLUSTRATIVE EXAMPLE

To provide an example of analyzing an NI trial using the question-based approach, the trial titled "Rivaroxaban for Thromboprophylaxis in Acutely Ill Medical Patients" (MAGELLAN), published in February, 2013 in the *New England Journal of Medicine*, and its published protocol, published earlier in February, 2011 in the *Journal of Thrombotic and Thrombolysis*, were chosen for this review (Cohen *et al*, 2011; Cohen *et al*, 2013). The MAGELLAN study aimed to test the NI of oral rivaroxaban compared with subcutaneous enoxaparin for prevention of thromboembolism (VTE) at 10 days. In addition, to examining the propensity of extended thromboprophylaxis duration for prolonged immobilized medical patients, it also tested the superiority of rivaroxaban over placebo following 10 days of enoxaparin for thromboprophylaxis at 35 days as a coprimary outcome. Table 5 in this review summarizes the analysis of the MAGELLAN.

Table 5. Illustrative example using Magellan

Section/topic	**Questions to evaluate**	**MAGELLAN Study**
Title, Abstract, and Introduction		
Title and Abstract	Q1. Are the title and abstract appropriate to identify the study as a NI trial? Could it be discerned which experimental intervention is being compared against which active control?	Although the title of the trial did not specify its objective as, the abstract described the NI objective in its method, together with its superiority outcome (p. 513, MAGELLAN). Likewise, the abstract listed the experimental intervention as oral rivaroxaban and the active control as subcutaneous enoxaparin. To make it fully complete, the title could have been as follows: "Noninferiority and Superiority of Oral Rivaroxaban *vs*. Subcutaneous (s.c) Enoxaparin for Thrombophrophylaxis in Acutely Ill Medical Patients." The two treatment groups were identified in the abstract:
Introduction	Q2. Was the rationale for using NI design provided? Was the NI hypothesis stated?	Both the main article and its protocol did not explicitly explain why a NI design was used. However the authors did report certain ancillary benefits of oral rivaroxaban, with fast onset and offset, for prevention of VTE in the protocol (p. 408). The authors should have provide details about the NI design, such as the convenience of oral dosing for rivaroxaban compared with subcutaneous injection of enoxaparin and the usefulness of the new agent in patients with heparin-induced-thrombocytopenia (HIT). The rationale would give the readers a better understanding of rivaroxaban in VTE prophylaxis for acutely ill medical patients.

Table 5. Illustrative example using Magellan (Continued)

Section/topic	Questions to evaluate	MAGELLAN Study
		The protocol discussed the NI hypothesis as the first primary efficacy outcome in its study objectives: "to determine whether oral rivaroxaban 10 mg once daily (od) for 10±4 days is non-inferior to s.c enoxaparin (40 mg od) for 10±4 days for the prevention of VTE in patients aged 40 years or older and hospitalized for a medical illness." It is worth noticing that the co-primary outcome of superiority at 35±4 days was also mentioned in the protocol.
Methods		
Participants	Q3. Were the participants recruited in the NI trial similar to those in trial(s) that established efficacy of the active control?	The inclusion criteria discussed in the protocol established the pool of study participants who were admitted for medical illnesses that predisposed them to VTE risk, including heart failure, active cancer, acute ischemic stroke, acute infection and inflammation, acute respiratory insufficiency, and many other risk factors for developing VTE (p. 410 protocol). Upon examination of the meta-analysis of Leizorovicz and Mismetti, the VTE-prone medical illnesses, except for active cancer and acute ischemic stroke, and VTE risk factors described in the results section of MAGELLAN were similar to previous studies that established the use of low-molecular-weight heparin (LMWH) for prevention of VTE, particularly the MEDENOX and PREVENT trials (p. 518, MAGELLAN) (Leizorovicz and Mismetti, 2004). Active cancer and acute ischemic stroke were included in a different trial, EXCLAIM, which studied extended duration of subcutaneous enoxaparin compared with standard duration of subcutaneous enoxaparin (Hull *et al*, 2010)

Table 5. Illustrative example using Magellan (Continued)

Section/topic	Questions to evaluate	MAGELLAN Study
Interventions	Q. Was the active control appropriately chosen? Was the active control in the NI trial identical or similar to that of historical trial(s) that established its efficacy? Q. Were assigned treatments carried out appropriately?	The active control arm was standard duration of subcutaneous enoxaparin 40 mg once daily for 10±4 days for test of NI, followed by placebo until day 35±4 days for test of superiority (p. 514, MAGELLAN). The standard dose of 40 mg and duration of two weeks of enoxaparin injection was the appropriate VTE prophylaxis at the time of the study as established in previous trials, and thus this satisfied the NI objective. However, the following placebo period of the enoxaparin injection arm gave the oral rivaroxaban arm an edge over the active control for the superiority objective. On the other hand, the experimental intervention, 10 mg oral rivaroxaban once daily up to 35±4 days, was used appropriately as compared to prophylactic dose used for other VTE indications. Assigned treatments were carried out according to the protocol, as depicted in the study flow chart (p. 517, MAGELLAN).
Outcomes	Q. Were the NI outcomes stated? Were primary and secondary outcomes based upon NI or superiority hypotheses? Were the primary outcomes of NI identical or similar to that of historical trial(s) that established efficacy for the active control?	The NI outcome of the MAGELLAN was a composite efficacy outcome measured at 10±4 days, including asymptomatic proximal deep vein thrombosis (DVT), symptomatic DVT (proximal and distal), nonfatal pulmonary embolism (PE), and VTE-related death (p. 514, MAGELLAN). The same composite outcome was used for superiority testing at 35±4 days. Other secondary outcomes and bleeding safety outcomes were discussed in the study and the protocol. The primary NI composite outcome was appropriate and similar to previous trials, and methods of measuring outcomes were done previously (e.g., ultrasonography, venography, thoracic spiral computed tomography, chest radiography, or pulmonary angiography).

Table 5. Illustrative example using Magellan (Continued)

Section/topic	Questions to evaluate	MAGELLAN Study
Sample size and NI margin	Q. Was the sample size calculated using NI criterion? Q. Was the NI margin specified a priori with rationale? Q. Was NI margin reported in absolute or relative scale?	The sample size was calculated based on both NI and superiority outcomes, and the estimate of a sample size of 2876 participants was based on a NI margin of 1.5 from the lower bound of 95% CI from a meta-analysis of three trials comparing placebo with active control in medically ill patients (p 515, MAGELLAN; p. 413, protocol). The NI margin would mean that oral rivaroxaban is considered noninferior to subcutaneous enoxaparin if rivaroxaban efficacy were less effective by 50% or less on the relative scale. Further investigation revealed that this 1.5 NI margin had not applied any discounting or clinical adjustment for to avoid loss of efficacy of the active control (Leizorovicz and Mismetti, 2004). Thus the NI margin would be too large, and accepting NI at this margin would risk losing all efficacy of enoxaparin and thus rivaroxaban would not fare better than placebo. Recalculation of the NI margin using $dmax = 1.5$, and $f = 0.5$ (preserved 50% efficacy) would yield a NI margin of 1.22, following the point estimate method. This new NI margin would attempt to preserve 50% efficacy of enoxaparin compared to placebo in the new trial. The NI margin was reported in relative scale, which was the preferable method of reporting.
Statistical Analysis	Q. Was an intention-to-treat (ITT) or per-protocol (PP) analysis used? Was missing data appropriately handled? Q. Was a one-sided or two-sided confidence interval approach used?	Though the study described using both PP and modified ITT (mITT) (received at least one dose of study medication) for the NI objective and superiority objective, the authors reported only PP analysis for the NI outcome at day 10 and only mITT for the superiority objective at day 35. It would be more appropriate to compare both mITT and PP analyses for their NI outcomes. Missing data was handled with a probability approach for binary composite outcome (p. 413, protocol); however, the authors failed to report the results of such data in the main article. One-sided test set at 0.025 significance level was used appropriately for the efficacy outcomes, of both NI and superiority. A two-sided test set at 0.05 significance level was used for safety outcome.

Table 5. Illustrative example using Magellan (Continued)

Section/topic	Questions to evaluate	MAGELLAN Study
Results and Application to Patient Care		
Participant characteristics	Q. Was prognostic balance between experimental and control groups maintained before, during, and at completion of the trial?	Upon enrollment and randomization, baseline characteristics were well balanced between two groups (p. 518, MAGELLAN). Nevertheless, there was no report on maintaining balance during and at completion of the trial. Effect of dropouts on balance of participants' characteristics was not assessed. A more appropriate report of these data would include patients' characteristics in each analysis group to assess for balance between groups at each stage of analysis.
Outcomes and Estimation	Q. Were the study results of NI outcomes presented in relation to confidence intervals and NI margin? Q. Were sensitivity analyses done to test for robustness of results?	Primary NI outcome result was presented in table 2 of the main article, with relative risk and 95% CI 0.97 (0.71 - 1.31), well below the NI margin of 1.5 (p-value of 0.003 for NI using PP approach) (p. 519, MAGELLAN). A graphic presentation of the 95% CI in relation with the NI margin would be preferred. NI of oral rivaroxaban to enoxaparin injection was confirmed, and the sequential testing for co-primary outcome of superiority was carried out. Secondary outcome results were presented in table 3 of the main article. Although the protocol called for sensitivity analyses to be performed, there was no report of sensitivity analyses in the article (p. 413, protocol).
Discussion	Q. Were the results interpreted in relation to the NI hypothesis? Was conclusion of NI warranted? Q. Was the effect of the active control preserved? Q. If a superiority conclusion was drawn for the outcomes for which NI was hypothesized, did the authors provide justification for switching?	Because the 95% CI for the composite outcome at day 10$\pm$4 (the NI outcome) lied below the prespecified NI margin 1.5, the NI of oral rivaroxaban compared with enoxaparin injection was confirmed by the authors. However, as mentioned before, the NI margin of 1.5 was excessive, and thus the NI conclusion was not very well defended. The effect of the active control was somewhat maintained in the study, with enoxaparin group event rate at 2.7% slightly above the calculated 2.2%. Constancy assumption with regards to active control choice and dosing and participant selection had been adequately maintained, but there was neither a report using mITT for the NI outcome nor a report of sensitivity analysis, so the conclusion of NI was not robust. Superiority outcome at day 35$\pm$4 was tested and confirmed after the NI outcome has been confirmed,

Table 5. Illustrative example using Magellan (Continued)

Section/topic	Questions to evaluate	MAGELLAN Study
		but in view of the non-conclusiveness of the NI result, superiority results should not be evaluated.
Applicability to patient care	Q. Are my patients similar to the study participants? Q. Are the benefits of using noninferior intervention worth the trade-off of effectiveness from not using standard treatment? Is the loss of efficacy acceptable to my patients?	According to study results, a typical patient who would benefit from the study result should be at risk for developing VTE if they are admitted for medical illnesses that predispose them to have VTE and have at least one additional risk factors for developing VTE (with the exception of acute heart failure, acute ischemic stroke with leg paralysis, and active cancer which do not require additional VTE risk factors). Furthermore, they are likely of advanced age (71 years), likely to be overweight, and likely to be white. These are some considerations to be made when selecting the NI therapy for a patient. Recall that the 95% CI of 0.71 - 1.31 of the NI outcome at day 10$\pm$4 would mean that there is a possibility that oral rivaroxaban could be 31% less effective compared to subcutaneous enoxaparin. Acceptance of this 31% loss of efficacy needs to be balanced against the benefit of using rivaroxaban instead of enoxaparin, namely convenience of oral formulations and usability in patients with history of heparin-induced thrombocytopenia (HIT). Cost is definitely not an ancillary benefit of rivaroxaban (~$410 for 30 day-supply of 10 mg Xarelto) compared with generic enoxaparin (~$180 for 30 day-supply of 40 mg) (Micromedex, 2018). Furthermore, safety results indicated that rivaroxaban increased bleeding risk at day 10 compared with standard duration enoxaparin, with a significant relative risk of 2.3 (95% CI 1.63 - 3.17), absolute risk difference (ARD) of 1.6%, and number-needed to harm (NNH) of 63. At day 35, bleeding risk, relative risk was at 2.5 (95% CI 1.85 - 3.25), ARD of 2.4%, and NNH of 42. Given its efficacy loss of 31%, its lack of substantial ancillary benefits, and its safety issues, use of oral rivaroxaban for prevention of VTE in acutely ill medical patients should be limited only to those who cannot use enoxaparin injection and have a lower bleeding risk.

CONCLUSION

Analyzing an NI trial requires a thorough knowledge of the NI design and its related statistical concepts. Unlike conventional superiority trials, NI trials are based on historical evidence of effectiveness for validation, and thus designing an NI trial must be based on sound constancy assumptions and careful selection of a NI margin, besides other critical requirements of a good quality clinical study. Quality of NI trials and their reporting, unfortunately, stand in need of improvement, so pharmacists and health care practitioners should be able to critically evaluate the evidence presented within any NI trials and make evidence-based decision when applying the NI results to patient care.

REFERENCES

Agency for Healthcare Research and Quality (AHRQ). EPC Working Group. 2012. Assessing Equivalence and Non-Inferiority. AHRQ Publication No. 12-EHC045-EF. Retrieved from https://effectivehealthcare.ahrq.gov /sites/default/files/pdf/equivalence_research.pdf

Althunia TA, de Boer A, Klungel OH, Insani WN, Groenwold RHH. 2017. Methods of defining the non-inferiority margin in randomized double-blind controlled trials: a systematic review. *Trials*. 18:107. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5341347/pdf/13063_2017_Article_1859.pdf

Bryant PJ, McQueen CE. 2018. Chapter 5: Literature Evaluation II: Beyond the Basics. In: Malone PM, Malone MJ, Park SK. eds. *Drug Information: A Guide for Pharmacists*. 6th ed. New York, NY: McGraw-Hill.

Campbell-Scherer D. 2017. Reflections on using non-inferiority randomised placebo controlled trials in assessing cardiovascular safety of new agents for treatment of type 2 diabetes. *Evid Based Med.* 22(2): 54-56.

Cohen AT, Spiro TE, Buller HR, Haskell L, Hu D, Hull R, Mebazaa A, Merli G, Schellong S, Spyropoulos AC, Tapson V (MAGELLAN Investigators). 2013. Rivaroxaban for Thromboprophylaxis in Acutely Ill Medical Patients. *N Engl J Med.* 368(6):513-523.

Cohen AT, Spiro TE, Buller HR, Haskell L, Hu D, Hull R, Mebazaa A, Merli G, Schellong S, Spyropoulos AC, Tapson V. 2011. Extended-duration rivaroxaban thromboprophylaxis in acutely ill medical patients: MAGELLAN study protocol. *J Thromb Thrombolysis.* 31:407-416.

European Medicines Agenct (EMA). 2006. Committee for Medicinal Products for Human Use (CHMP): Guideline on the choice of the non-inferiority margin. Retrieved from http://www.ema.europa.eu/docs/en_GB/ document_library/Scientific_guideline/2009/09/WC500003636.pdf

EMA. 2001. Committee for Proprietary Medicinal Products (CPMP). Points to Consider on switching between superiority and non-inferiority. *J Clin Pharmacol*. 52:223-228

Head SJ, Kaul S, Bogers AJJC, Kappetein AP. 2012. Non-inferiority study design: lessons to be learned from cardiovascular trials. *European Heart Journa*l. 33:1318-1324.

Henanff AL, Giraudeau B, Baron G, Ravaud P. 2006. Quality of Reporting of Noninferiority and Equivalence Randomized Trials. *JAMA.* 295(10):1147-1151.

Hull RD, Schellong SM, Tapson VF, Monreal M, Samama M, Nicol P, Vicaut E, Turpie AGG, Yusen RD, EXCLAIM study. 2010. Extended-Duration Venous Thromboembolism Prophylaxis in Acutely Ill Medical Patients With Recently Reduced Mobility: A Randomized Trial. *Ann Intern Med*. 153(1):8-18.

Kaji AH, Lewis RJ. 2015. Noninferiority Trials: Is a New Treatment Almost as Effective as Another? *JAMA.* 313(23):2371-2372.

Kaul S, Diamond G. 2006. Good Enough: A Primer on the Analysis and Interpretation of Noninferiority Trials. *Ann Intern Med*. 145(1):62-69

Leizorovicz A, Mismetti P. 2004. Preventing Venous Thromboembolism in Medical Patients. *Circulation*. 110[suppl IV]:IV-13 - IV-19.

Mauri L, D'Agostino RB, Sr. 2017. Challenges in the Design and Interpretation of Noninferiority Trials. In: Drazen MJ, Harrington PD, McMurray JVJ, Ware HJ, Woodcock J. eds. *The Changing Face of Clinical Trials. N Engl J Med.* 377(14):1357-1367.

Micromedex (IBM Corporation). 2018. Product Lookup - Red Book: enoxaparin sodium and rivaroxaban. Accessed May 7, 2018.

Mulla SM, Scott IA, Jackevicius CA, You JJ, Guyatt GH. 2012. How to use a Noninferiority Trial: Users' Guides to the Medical Literature. *JAMA.* 308(24):2605-2611.

Palmas W. 2017. The CONSORT guidelines for noninferiority trials should be updated to go beyond the absolute risk difference. *Journal of Clinical Epidemiology*. 83:6-7

Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG. CONSORT group. 2012. Reporting of Noninferiority and Equivalence Randomized Trials: Extension of the CONSORT 2010 Statement. *JAMA*. 308(24):2594-2604.

Rehal S, Morris TP, Fielding K, Carpenter JR, Phillips PPJ. 2016. Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. *BMJ Open*. 6: e012594. doi:10.1136/bmjopen-2016-012594

Schumi J, Wittes JT. 2011. Through the looking glass: understanding non-inferiority. *Trials*. 12: 106. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3113981/pdf/1745-6215-12-106.pdf

Suda KJ, Hurley AM, Mckibbin T, Moroney SEM. 2011. Publication of Noninferiority Clinical Trials: Changes Over a 20-Year Interval. *Pharmacotherapy*. 31(9):833-839.

United States Food and Drug Administration (U.S. FDA). 2008. Guidance for industry: Diabetes Mellitus - Evaluating Cardiovascular Risk in New Antidiabetic Therapies to Treat Type 2 Diabetes. Retrieved from https://www.fda.gov/downloads/Drugs/Guidances/ucm071627.pdf

United States Food and Drug Administration (U.S. FDA) 2016. Non-Inferiority Clinical Trials to Establish Effectiveness: Guidance for Industry. Retrieved from https://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf

Vermeulen L. 2011. Gain in Popularity of Noninferiority Trial Design: Caveat Lector. *Pharmacotherapy.* 31(9):831-832.