

# Critically appraising noninferiority randomized controlled trials: a primer for emergency physicians

Mohammad Al Deeb, MD\*<sup>†</sup>; Aftab Azad, MBBS\*<sup>‡</sup>; David Barbic, MD, MSc<sup>§</sup>

## ABSTRACT

Noninferiority (NI) trials aim to show that a new treatment or drug is not inferior to a standard, accepted treatment. The rapid proliferation of NI trials within the literature makes it imperative for emergency physicians to be able to read, interpret, and appraise critically this type of research study. Using several emergency medicine examples from the recent literature, this article outlines the key differences between traditional, superiority randomized controlled trials and NI trials. We summarize four important points that an emergency physician should consider when critically appraising an NI trial: 1) Does the new treatment have tangible benefits over the standard treatment? 2) Was the choice of the NI margin appropriate? 3) Was the effect of the standard treatment preserved? Does the trial have assay sensitivity? and 4) What type of analysis strategy was employed: intention-to-treat (ITT) or per protocol (PP)?

## RÉSUMÉ

Les essais de non-infériorité (ENI) visent à démontrer qu'un nouveau traitement ou un nouveau médicament n'est pas inférieur au traitement de référence, reconnu. La prolifération des ENI dans la documentation rend impérative la capacité des urgentologues à comprendre, à interpréter et à évaluer d'un œil critique ce type de recherche. Ainsi, l'article fera ressortir, à l'aide de plusieurs exemples tirés de la documentation récente en médecine d'urgence, les principales différences qui existent entre le modèle classique d'essai comparatif, hasardisé, de supériorité, et le modèle d'ENI. Seront résumés quatre points importants qu'un urgentologue devrait envisager lorsqu'il évalue de façon critique un ENI: 1) Le nouveau traitement offre-t-il des avantages tangibles comparativement au traitement habituel? 2) Le choix de l'intervalle de non-infériorité est-il pertinent? 3) L'effet du traitement de référence est-il conservé? L'essai est-il sensible? 4) Quel type de stratégie d'analyse a été employé: selon l'intention de traiter ou selon le protocole?

**Keywords:** emergency medicine, methodology, noninferiority, randomized controlled trial, review

## CASE SCENARIO

You are an emergency physician practicing in a busy Canadian emergency department (ED), and you have been asked by your department head to review the literature regarding the implementation of the latest Surviving Sepsis Campaign guidelines<sup>1</sup> and early goal-directed therapy (EGDT)<sup>2</sup> in your ED. Although currently in use by some clinicians, there has been moderate uptake because of the perceived resource challenges of widespread central venous catheterization and central venous oxygen saturation (ScvO<sub>2</sub>) monitoring.<sup>3,4</sup> Your department head recently heard a podcast about lactate clearance in EGDT and would like to know if this is an acceptable alternative to implement as part of your department's EGDT bundle.<sup>5</sup>

## INTRODUCTION

The objective of noninferiority (NI) trials (sometimes referred to as active control equivalence trials) is to show that a new treatment or drug is not inferior to a standard, accepted treatment.<sup>6</sup> NI trials are not designed to show that the new treatment is equivalent to the standard treatment. NI trials are intended to show whether a new treatment has at least as much efficacy as the standard or is worse by an amount less than a prespecified NI margin.<sup>7</sup> The premise of the new treatment is that it offers some other advantage,

From the \*Department of Emergency Medicine, McGill University Health Centre, Montreal, QC; †Department of Emergency Medicine, King Abdulaziz Medical City, Riyadh, Saudi Arabia; ‡Department of Emergency Medicine, Hamad Medical Corporation, Doha, Qatar; and Department of Emergency Medicine, Sunnybrook Health Sciences Centre, Toronto, ON.

**Correspondence to:** Dr. David Barbic, Department of Emergency Medicine, Sunnybrook Health Sciences Centre, 2075 Bayview Avenue, Toronto, ON M4N 3N5; david.barbic@gmail.com.

This article has been peer reviewed.

© Canadian Association of Emergency Physicians

CJEM 2015;17(3):231-236

DOI 10.2310/8000.2014.141405

DECKER



CJEM • JCMU

2015;17(3) 231

such as greater availability, reduced cost, less invasiveness, fewer side effects, or greater ease of administration (once-daily dosing rather than two or more doses).<sup>7</sup> Other differences between superiority randomized controlled trials (RCTs) and NI trials include the initial study hypotheses, trial design, data analysis, and interpretation of results.<sup>7</sup>

## **DIFFERENCES BETWEEN SUPERIORITY AND NI TRIALS**

### ***Study hypotheses***

In a superiority RCT, the null hypothesis is that treatments are equally effective, and the alternative hypothesis is that they differ. Type I errors occur when a treatment effect is falsely observed when, in fact, there is none. Type II errors occur when a treatment effect is not detected when one truly exists.<sup>7</sup> In NI trials, the inverse is true, and the null and alternative hypotheses are reversed; Type I errors occur with the false acceptance of an inferior new treatment, and Type II errors occur with the false rejection of a truly noninferior treatment.<sup>7</sup>

### ***Trial design***

NI trials require that the efficacy of the reference treatment is well established or is in widespread use, making the use of placebo or an untreated control group unethical. NI trials should aim to recruit similar patients, have similar inclusion and exclusion criteria, and use outcome measures similar to those of trials that established the efficacy of the reference treatment.<sup>7</sup> An NI margin must be specified a priori to calculate a feasible sample size to conduct the study. There are several approaches to set an NI margin,<sup>7-11</sup> and controversy regarding the best method persists.<sup>6,12,13</sup> This is discussed in more detail in the Critical Appraisal of an NI Trial section.

### ***Data analysis***

In superiority RCTs, intention-to-treat (ITT) analysis is widely recommended and often leads to smaller treatment effects than if all patients adhered to the treatment protocol.<sup>14</sup> The analysis of NI trials requires careful consideration of both the ITT and per protocol (PP) strategies to avoid erroneous Type I and Type II errors.<sup>9,11</sup> Superiority RCTs employ a two-tailed (two-sided)

hypothesis test, whereas NI trials employ a one-tailed (one-sided) hypothesis test.<sup>15,16</sup>

### ***Interpretation***

Interpreting an NI trial's results depends on where the confidence interval for the treatment effect lies relative to both the NI margin and the null effect.<sup>7</sup> An important difference between superiority and NI trial studies, in terms of inference, is that in superiority studies, a positive result is self-explanatory. Apart from ensuring that the study was conducted adequately, one can make inferences and conclusions directly from the results. NI trials require additional information outside the study to be able to completely and adequately arrive at valid conclusions. (This is discussed in detail below.) It is important to note that it is inappropriate to claim NI post hoc from a superiority RCT unless clearly stated in the trial protocol with a predefined margin of equivalence.<sup>17</sup>

How common are NI trials in the literature? A quick search of *MEDLINE* reveals over 589 such studies published within the last 5 years. Why would an investigator choose to conduct an NI trial as opposed to a superiority RCT? Very often, the new treatment offers some other advantage compared to the standard treatment, such as greater availability, reduced cost, less invasiveness, fewer side effects, and greater ease of administration (once-daily dosing rather than two or more doses), but is not expected to be superior to the standard treatment. In this setting, comparing the new treatment to the standard treatment by designing a superiority trial might prove to be futile or, at best, very difficult. This trial would need a very large sample size given that the new treatment is not expected to be superior. An NI design in this setting would be more appropriate because a prespecified NI margin would ensure that a feasible sample size could be obtained. The recent Rapid Anticonvulsant Medications Prior to Arrival Trial (RAMPART) examining intramuscular (IM) midazolam versus intravenous (IV) lorazepam for the termination of prehospital status epilepticus reported several potential benefits of the new treatment (midazolam).<sup>18</sup> These included easier and more reliable administration and a longer shelf-life than lorazepam when not refrigerated.<sup>18</sup> Other recent examples relevant to emergency medicine are the trials designed to show NI between the novel anticoagulants and

warfarin in the treatment of venous thromboembolism.<sup>19,20</sup> These NI trials have led to guidelines advocating rivaroxaban as an alternative first-line option in the treatment of acute deep vein thrombosis and pulmonary embolism.<sup>21</sup>

Returning to the case scenario, after performing a search of the literature, you identify the 2010 NI trial by Jones and colleagues.<sup>5</sup> The stated objective of this study was to test the hypothesis of NI between lactate clearance and ScvO<sub>2</sub> as goals of early sepsis resuscitation. You endeavour to interpret the results and critically appraise this article.

**INTERPRETATION OF THE RESULTS OF AN NI TRIAL**

In this section, we discuss the potential outcomes and the means to interpret the results of an NI trial. Consider the following examples in Figure 1. The line labeled “delta” represents the NI threshold compared to the standard treatment. The area between “delta” and zero risk difference or relative risk represents the NI margin. As in Figure 1:

1. If the result is as in study A, we can conclude that the novel treatment is noninferior to the standard treatment because the lower limit of the 95% CI does not cross the NI margin. We can conclude superiority because the lower limit of the 95% CI does not cross the zero mark for risk difference or relative risk.

2. If the study yields result B, we can conclude that the novel treatment is noninferior to the standard but are unable to comment on superiority.
3. In study C, the results are indeterminate, and we are not able to say with confidence whether the novel treatment is noninferior or truly inferior because the lower limit of the 95% CI falls below the NI margin.
4. In study D, we are forced to conclude that the novel treatment is inferior to the standard.

Returning to our case scenario and the NI trial by Jones and colleagues,<sup>5</sup> for the primary study end point of in-hospital mortality, the group followed with lactate clearance demonstrated NI outcomes comparable to those followed with ScvO<sub>2</sub> monitoring.<sup>5</sup> The result demonstrated by this trial is similar to scenario B in Figure 1.

**CRITICAL APPRAISAL OF AN NI TRIAL**

NI studies are appraised in a manner similar to that of superiority RCTs, with several important additions and caveats.<sup>9</sup> To critically appraise an NI trial, additional information from outside the study may be required to come up with valid conclusions. This is one of the major difficulties and limitations of NI trials. Similar to appraising a meta-analysis, it is often necessary to examine the articles that formed the foundation for the efficacy of the standard treatment and set the choice of the NI margin.

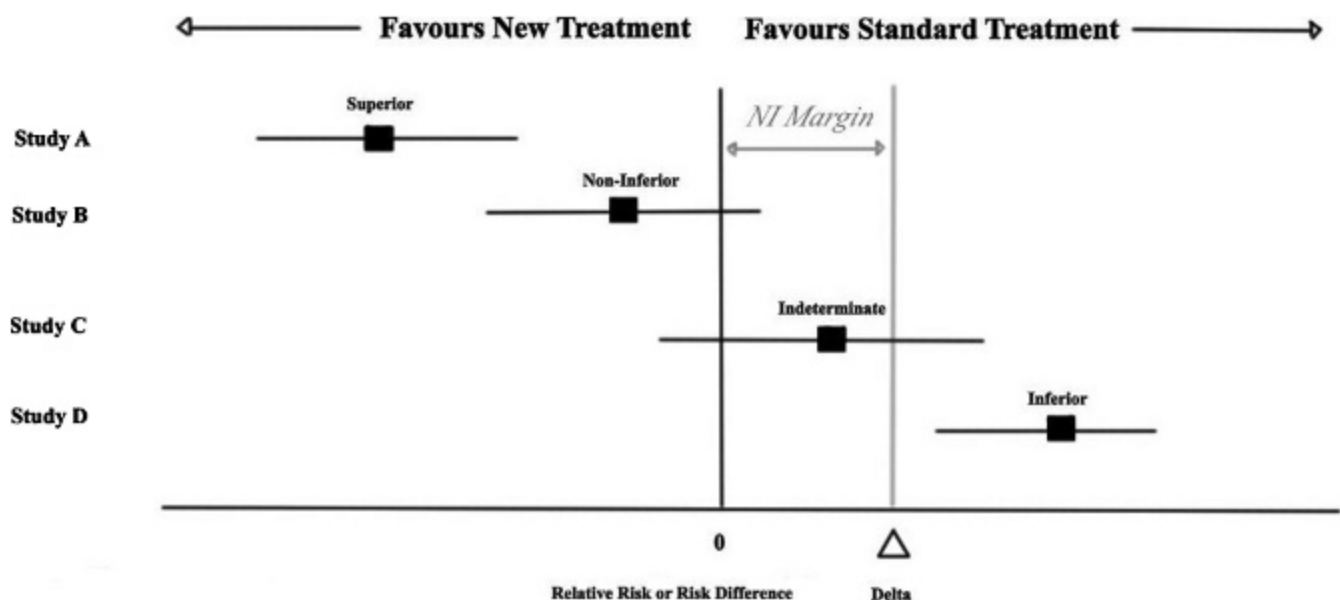


Figure 1. Possible outcomes of a noninferiority trial.

### **1. Does the new treatment have tangible benefits over the standard control?**

When critically appraising an NI trial, one needs to ensure that the new treatment has some tangible advantages over the standard treatment that will compensate for its potential inferiority and not merely be a “me too” drug (e.g., greater availability, reduced cost, less invasiveness, fewer side effects, greater ease of administration [once-daily dosing rather than two or more doses]).<sup>7</sup> Otherwise, there might be no reason to test the new treatment. The authors need to clearly and explicitly explain the rationale for choosing the new treatment.

### **2. Was the choice of the NI margin appropriate?**

There is no clear consensus on the ideal method for determining an NI margin. Experts have recommended sound statistical reasoning and good clinical judgment in setting an NI margin.<sup>6,13,22,23</sup> Recent draft guidance to industry from the US Food and Drug Administration (FDA) has provided some clarity,<sup>11</sup> but even this approach is not without its critics.<sup>12</sup> We outline several of the common approaches employed to choose an NI margin.

The suggested method from the FDA focuses on considering the smallest plausible treatment benefit of the standard treatment compared to placebo. Investigators usually look at previous superiority RCTs and identify the effect size of the standard treatment compared to placebo. These are the trials in which the control treatment established its superiority and efficacy. Investigators would then take the lower limit of the 95% CI of the effect size of the standard treatment relative to placebo and take 50% of that as the NI margin.<sup>9,11</sup> The logic is that one should at least preserve a proportion of the effect of the standard treatment. Choosing a higher percentage to be preserved (e.g., 67% or 90%) results in a stricter or more conservative NI margin.<sup>6</sup>

Another approach is to ask clinicians or patients to consider what degree of efficacy they would be willing to sacrifice in exchange for the potential benefits offered by the new treatment.<sup>10</sup> A Delphi approach with clinical experts and their knowledge of existing treatment options and the underlying disease may provide insight into an NI margin. Patient groups may provide insight into trade-offs that future patients might be willing to accept for a treatment with benefits, such as an improved dosing schedule or fewer side effects.<sup>10</sup>

A third method, the synthesis method, does not require a specific margin or active control effect to be specified.<sup>10</sup> The test of the NI hypothesis is based on a combination of the estimate and standard errors for the comparison of the standard versus placebo in the original superiority RCT and the estimate and standard error of the new treatment with the standard in the NI trial.<sup>10</sup>

Returning to our case scenario, Jones and colleagues chose an NI margin of  $-10\%$ .<sup>5</sup> This represented a two-thirds proportion of the active comparator's group, the established superiority in the original EGDT trial.<sup>2</sup> Evidence exists that mortality rates for severe sepsis can be expected to differ by up to 10% between both state-of-the-art medical facilities and regions of the United States.<sup>5</sup> This method for setting the NI margin is similar to that suggested by the recent FDA document.<sup>11</sup>

### **3. Was the effect of the standard treatment preserved? Does the trial have assay sensitivity?**

It is important to consider whether the treatment effect observed for the standard treatment in the NI trial is similar to the previous superiority RCTs with a placebo control.<sup>9</sup> Suboptimal application of the standard treatment, enrolling patients less likely to be adherent or responsive to the standard treatment, enrolling patients at low risk for the effectiveness outcome, and administering treatment via a suboptimal route (e.g., oral versus intravenously) may all contribute to a decreased observed treatment effect. It is important to examine the original superiority RCTs, which determined the effectiveness of the standard treatment compared to placebo.<sup>9,10</sup> This will ensure that the current NI trial has assay sensitivity.

The assay sensitivity of a study can be explained as the ability of the study to detect a difference between treatments for a particular outcome, if one truly exists.<sup>8</sup> A hypothetical example might be an NI trial examining IM midazolam and IV lorazepam for status epilepticus. The trial finds that the treatment effects are exactly the same: 5% of patients had their seizures terminated in both groups. One would conclude that IM midazolam is noninferior to IV lorazepam. If it is discovered that there was an error in the study, and all the medications used were expired, then one would need to revisit the conclusions of the study. Both study drugs were equivalent because neither worked, and one would say that the study did not have assay sensitivity. This is

a radical example, and in reality, subtle things might decrease the assay sensitivity: poor conduct of the study, an inordinate number of dropouts, loss to follow-up, and poor adherence to the study protocol. The presence of assay sensitivity is much more important to demonstrate in NI trials compared to superiority RCTs because its absence may result in Type I errors and the erroneous acceptance of a noninferior new treatment.<sup>7</sup>

Returning to our case scenario, the original Rivers and colleagues study showed an in-hospital mortality rate of 46.5% for patients receiving standard therapy and 30.5% for patients enrolled in the EGDT group.<sup>2</sup> The study by Jones and colleagues demonstrated an in-hospital mortality of 23% for the similar (ScVO<sub>2</sub>) group.<sup>5</sup> There has been vigorous debate as to whether the NI trial by Jones and colleagues included patients as sick as those in the original EGDT study<sup>24</sup> or whether the original EGDT study was truly representative of the mortality rate from severe sepsis and septic shock.<sup>25</sup> It is currently challenging to say with certainty whether the NI trial by Jones and colleagues possesses assay sensitivity.

#### **4. What type of analysis strategy was employed? ITT or PP?**

The type of analysis employed hinges on the previous concept of assay sensitivity. Recall that assay sensitivity decreases when there are validity issues with a trial: high dropout rates, poor adherence to protocol, or poor adherence to inclusion criteria. By analyzing the results as ITT, we include in our analysis all the subjects who have been randomized, regardless of whether they adhered to the protocol. Although this might decrease the assay sensitivity, this is an appropriate analytical method in superiority RCTs that will ensure the adequacy of the randomization process and will reflect reality.<sup>14</sup> By employing this approach, we are being conservative in accepting any results from a superiority RCT that will potentially change practice. Analysis of an NI trial with the ITT approach alone runs the risk of decreasing the assay sensitivity of the trial and the acceptance of a Type I error. Consequently, the suggested method of NI trial analysis is examination of the data from both the ITT and PP perspectives.<sup>11,26</sup> The FDA cites “significant concerns with the possibility of informative censoring” in an “as-treated” or PP analysis and states that discrepancies between the two

types of analyses will require “close examination.”<sup>10,11</sup> The results of an NI trial are valid and convincing if they are positive by both analytical methods.

Returning to our case scenario, the NI trial by Jones and colleagues employed both ITT and PP analytical approaches,<sup>5</sup> and by both approaches, lactate clearance was shown to be statistically noninferior to ScVO<sub>2</sub> measurement for the primary outcome, in-hospital mortality, in the resuscitation of septic patients.

#### **SUMMARY**

NI trials are a powerful tool for assessing outcomes relevant to emergency medicine clinical practice. Multiple recent examples from the literature highlight the need for emergency physicians to understand the subtle differences between superiority and NI trials. Critical appraisal of these studies requires consideration of four main points: 1) Does the new treatment have tangible benefits over the standard treatment? 2) Was the choice of the NI margin appropriate? 3) Was the effect of the standard treatment preserved? Does the trial have assay sensitivity? 4) What type of analysis strategy was employed? ITT or PP? We hope that emergency physicians feel better equipped to read and critically appraise these studies with the aid of this review article.

**Competing interests:** None declared.

#### **REFERENCES**

1. Dellinger RP, Levy MM, Rhodes A, et al. Surviving Sepsis Campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med* 2013; 41:580-637, doi:10.1097/CCM.0b013e31827e83af.
2. Rivers E, Nguyen B, Havstad S, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 2001;345:1368-77, doi:10.1056/NEJMoa010307.
3. Jones A, Kline JA. Use of goal-directed therapy for severe sepsis and septic shock in academic emergency departments. *Crit Care Med* 2005;33:1888-9, doi:10.1097/01.CCM.0000166872.78449.B1.
4. Jones AE, Shapiro NI, Roshon M. Implementing early goal-directed therapy in the emergency setting: the challenges and experiences of translating research innovations into clinical reality in academic and community settings. *Acad Emerg Med* 2007;14:1072-8, doi:10.1111/j.1553-2712.2007.tb02391.x.
5. Jones AE, Shapiro NI, Trzeciak S, et al. Lactate clearance vs central venous oxygen saturation as goals of early sepsis therapy: a randomized clinical trial. *JAMA* 2010;303:739-46, doi:10.1001/jama.2010.158.
6. Wange G, Roes KC, de Boer A, et al. The challenges of determining noninferiority margins: a case study of non

- inferiority randomized controlled trials of novel anticoagulants. *CMAJ* 2012;185:222-7, doi:[10.1503/cmaj.120142](https://doi.org/10.1503/cmaj.120142).
7. Piaggio G, Elbourne DR, Altman DG, et al. Reporting of noninferiority and equivalence randomized trials. *JAMA* 2006;295:1152-60, doi:[10.1001/jama.295.10.1152](https://doi.org/10.1001/jama.295.10.1152).
  8. D'Agostino RB, Massaro JM, Sullivan LM. Non-inferiority trials: design concepts and issues - the encounters of academic consultants in statistics. *Stat Med* 2003;22:169-86, doi:[10.1002/sim.1425](https://doi.org/10.1002/sim.1425).
  9. Mulla SM, Scott IA, Jackevicius CA, et al. How to use a noninferiority trial: users' guides to the medical literature. *JAMA* 2012;308:2605-11, doi:[10.1001/2012.jama.11235](https://doi.org/10.1001/2012.jama.11235).
  10. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. *Trials* 2011;12(1):106, doi:[10.1186/1745-6215-12-106](https://doi.org/10.1186/1745-6215-12-106).
  11. Center for Biologics Evaluation and Research, US Food and Drug Administration. *Guidance for industry non-inferiority clinical trials*. February 24, 2010. Silver Spring (MD): US Food and Drug Administration; 2010.
  12. Huitfeldt B, Hummel J. The draft FDA guideline on non-inferiority clinical trials: a critical review from European pharmaceutical industry statisticians. *Pharm Stat* 2011;10:414-9, doi:[10.1002/pst.508](https://doi.org/10.1002/pst.508).
  13. Wangge G, de Boer A, Klungel OH, et al. Expert-opinion on non-inferiority margin: a case study of oral anti-coagulant agents for prophylaxis of venous thromboembolic events after orthopaedic surgery. *Thromb Res* 2013;131:368-71, doi:[10.1016/j.thromres.2013.01.013](https://doi.org/10.1016/j.thromres.2013.01.013).
  14. Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94, doi:[10.7326/0003-4819-134-8-200104170-00012](https://doi.org/10.7326/0003-4819-134-8-200104170-00012).
  15. Tamayo-Sarver JH. Advanced statistics: how to determine whether your intervention is different, at least as effective as, or equivalent: a basic introduction. *Acad Emerg Med* 2005;12:536-42, doi:[10.1111/j.1553-2712.2005.tb00897.x](https://doi.org/10.1111/j.1553-2712.2005.tb00897.x).
  16. ICH Expert Working Group. Statistical principles for clinical trials. ICH harmonised tripartite guideline. *Statistics in Medicine* 1999;18:1905-42.
  17. Ferguson JJ, Califf RM, Antman EM, et al. Enoxaparin vs unfractionated heparin in high-risk patients with non-ST-segment elevation acute coronary syndromes managed with an intended early invasive strategy: primary results of the SYNERGY randomized trial. *JAMA* 2004;292:45-54.
  18. Silbergleit R, Durkalski V, Lowenstein D, et al. Intramuscular versus intravenous therapy for prehospital status epilepticus. *N Engl J Med* 2012;366:591-600, doi:[10.1056/NEJMoa1107494](https://doi.org/10.1056/NEJMoa1107494).
  19. EINSTEIN Investigators. Oral rivaroxaban for symptomatic venous thromboembolism. *N Engl J Med* 2010;363:2499-510, doi:[10.1056/NEJMoa1007903](https://doi.org/10.1056/NEJMoa1007903).
  20. Schulman S, Kearon C, Kakkar AK, et al. Dabigatran versus warfarin in the treatment of acute venous thromboembolism. *N Engl J Med* 2013;361:2342-52, doi:[10.1056/NEJMoa0906598](https://doi.org/10.1056/NEJMoa0906598).
  21. Kearon C, Akl EA, Comerota AJ, et al. Antithrombotic therapy for VTE disease, Antithrombotic therapy and prevention of thrombosis, 9th edition. American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest* 2012;141(2 Suppl):e4195.
  22. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med* 2006;145:62-9, doi:[10.7326/0003-4819-145-1-200607040-00011](https://doi.org/10.7326/0003-4819-145-1-200607040-00011).
  23. Fleming TR. Current issues in non-inferiority trials. *Stat Med* 2008;27:317-32, doi:[10.1002/sim.2855](https://doi.org/10.1002/sim.2855).
  24. Rivers EP, Elkin R, Cannon CM. Counterpoint: should lactate clearance be substituted for central venous oxygen saturation as goals of early severe sepsis and septic shock therapy? No. *Chest* 2011;140:1408-13, doi:[10.1378/chest.11-2563](https://doi.org/10.1378/chest.11-2563).
  25. Jones AE. Point: should lactate clearance be substituted for central venous oxygen saturation as goals for early severe sepsis and septic shock therapy? Yes. *Chest* 2011;140:1406-7, doi:[10.1378/chest.11-2560](https://doi.org/10.1378/chest.11-2560).
  26. Committee for Proprietary Medicinal Products. *Points to consider on switching between superiority and non-inferiority*. London (UK): European Agency for the Evaluation of Medicinal Products; 2000.