

# How to Use a Noninferiority Trial

## Users' Guides to the Medical Literature

Sohail M. Mulla, BHSc

Ian A. Scott, MBBS, FRACP, MHA

Cynthia A. Jackevicius, PharmD, MSc

John J. You, MD, MSc

Gordon H. Guyatt, MD, MSc

### CLINICAL SCENARIO

You are an internist seeing a 51-year-old woman with severe osteoarthritis and limited mobility who presents with progressive dyspnea over a 3-day period. She is subjectively in distress, with a pulse rate of 105, a respiratory rate of 28 breaths per minute, and an arterial oxygen saturation of 85% while breathing room air. Aside from her arthritis, the physical examination is unremarkable, and the lower extremity examination shows no sign of deep venous thrombosis. A computed tomographic (CT) pulmonary angiogram reveals unequivocal clot in 2 lobar arteries.

Recently, you have been treating patients with deep venous thrombosis without hospital admission using low-molecular-weight heparin (LMWH) administration in the outpatient setting. You are less comfortable with this approach in the more dangerous setting of pulmonary embolism. You recall receiving, from the updating service to which you subscribe, a recent trial that addressed this issue. Before discussing the issue of inpatient vs outpatient treatment with your patient, you quickly retire to your office to review the article.<sup>1</sup> In doing so, you find that the trial tested for noninferiority and you wonder, as you begin to read the methods and results, if there are special issues you should consider when using this article to guide your clinical care.

See also page 2594.

Clinical investigators are increasingly testing treatments that have the primary benefit of decreased burden or harms relative to an existing standard. The goal of the resulting randomized trials—called noninferiority trials—is to establish that the novel treatment's effectiveness is not substantially less than the existing standard. Conclusions from these trials are, however, based on noninferiority thresholds specified by authors whose judgments may not coincide with those of patients and clinicians. This article highlights issues related to validity, interpretation, and applicability of results specific to noninferiority trials. Suboptimal administration of standard treatment or exclusive reliance on the analyze-as-randomized approach that is standard for conventional superiority trials may produce misleading results in noninferiority trials. Clinicians should judge whether the novel treatment's impact on effectiveness outcomes—the prime reason for wanting to prescribe it—is sufficiently close to that of standard treatment that they are comfortable substituting it for the existing standard. Trading off desirable and undesirable consequences is an individual decision: given the benefits of a novel treatment, some patients may perceive the uncertainty regarding a reduction in treatment effectiveness as acceptable while others may not.

*JAMA.* 2012;308(24):2605-2611

www.jama.com

### INTRODUCTION

Traditionally, randomized clinical trials (RCTs) have sought to ascertain whether a novel treatment is superior to standard treatment or placebo in improving quality of life or preventing morbid or mortal events—what we will refer to as effectiveness outcomes. In these superiority trials, the primary objective is to determine the magnitude of increased benefit of the novel intervention over standard therapy on effectiveness outcomes.

Recently, another paradigm has emerged that offers novel treatments not on the basis of superiority in effectiveness outcomes but instead because they reduce harms or other treatment burdens relative to standard treatment. In modern medicine, clinicians are fortunate to have many effective treatments; unfortunately, these

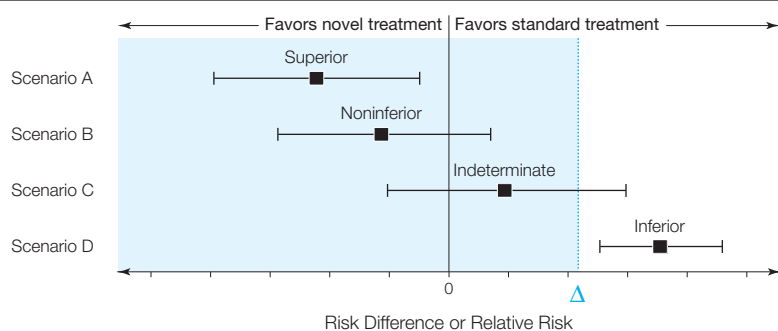
treatments are often associated with harms, inconvenience, or excessive cost. For these interventions, reducing treat-

**Author Affiliations:** Departments of Clinical Epidemiology and Biostatistics (Mr Mulla and Drs You and Guyatt) and Medicine (Drs You and Guyatt), Faculty of Health Sciences, McMaster University, Hamilton, Ontario, Canada; Department of Internal Medicine and Clinical Epidemiology, Princess Alexandra Hospital, Brisbane, Australia (Dr Scott); Institute for Clinical Evaluative Sciences, Toronto, Ontario, Canada (Drs Jackevicius and You); Department of Pharmacy Practice and Administration, College of Pharmacy, Western University of Health Sciences, Pomona, California (Dr Jackevicius); Institute of Health Policy, Management, and Evaluation, Faculty of Medicine, University of Toronto, (Dr Jackevicius); and University Health Network (Dr Jackevicius), Toronto, Ontario, Canada; and Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, California (Dr Jackevicius).

**Corresponding Author:** Gordon H. Guyatt, MD, MSc, Departments of Clinical Epidemiology and Biostatistics and Medicine, Faculty of Health Sciences, McMaster University, 1200 Main St W, Room 2C12, Hamilton, ON, L8N 3Z5 Canada (guyatt@mcmaster.ca).

**Users' Guides to the Medical Literature Section Editor:** Drummond Rennie, MD, Deputy Editor, *JAMA*.

**Figure 1.** Possible Outcome Scenarios in Noninferiority Trials



The blue dashed line labeled  $\Delta$  represents the noninferiority threshold or the maximum allowable excess of outcome events arising from the novel treatment compared with the standard treatment. The tinted area represents the noninferiority zone.

ment burden, including limitations and inconvenience, becomes a legitimate goal of innovative therapy.

In such instances, a question arises: can clinicians be confident that the novel treatment's impact on effectiveness outcomes—the prime reason for wanting to prescribe it—is sufficiently close to that of standard treatment that they are comfortable substituting it for the existing standard? In technical terms, is the novel treatment noninferior to the standard treatment?

Noninferiority trials provide an alternative to equivalence trials, which endeavor to establish that a novel treatment is neither better nor worse (beyond a specified margin) than the standard. In contrast, the noninferiority trialist is unconcerned if the novel treatment is better, so long as it is “not much worse.” Perhaps illustrating the limitations of the term, a noninferior treatment may thus be inferior, just not so inferior that it would cause concern. How much worse (ie, how much less effective) clinicians should consider acceptable will depend on the importance of the effectiveness outcome and the magnitude of the reduction in harms or burden achieved by the new treatment.

Consider how the concept of not much worse plays out for the patient presented in the scenario above. She may dislike spending time in the hospital and may strongly prefer treatment at home. But there may be risks

she incurs in choosing home management. Perhaps the care she would receive in the hospital would result in a lower risk of recurrent venous thromboembolism (VTE) and a lower risk of serious bleeding, which can complicate antithrombotic therapy. Would our patient be willing to incur the additional risk of VTE or serious bleeding that may be associated with home management? If so, what level of increased risk would she be willing to tolerate?

The example illustrates that, given that patients will choose the new treatment only if the risks are not much worse than the standard treatment, the critical issue in interpreting noninferiority trials is the choice of an acceptable threshold of not much worse. This noninferiority threshold (the dashed line labeled  $\Delta$  in FIGURE 1) is the maximum allowable excess of outcome events arising from the novel treatment compared with the standard treatment. When designing noninferiority trials, investigators set their own thresholds, typically using statistically based criteria (BOX 1). Sometimes, these standard approaches are not applicable, as was the case in the trial of inpatient vs outpatient treatment for pulmonary embolism (BOX 2).

When subsequently reviewing their results, if investigators find that the CI around the estimate for the difference in primary outcome events lies entirely below their chosen noninferiority threshold, they will claim noninferiority

(Figure 1, scenario B) or, in some instances, even superiority of the novel treatment (Figure 1, scenario A). If, on the other hand, the CI crosses the threshold, the trial has failed to establish noninferiority (Figure 1, scenario C). If the CI lies wholly above the noninferiority threshold, then the novel treatment is inferior to standard treatment (Figure 1, scenario D).

If noninferiority trials choose insufficiently stringent thresholds, they run the risk of concluding noninferiority when many patients might be unwilling to accept the novel treatment if they were informed of the largest possible increased risk (ie, decreased effectiveness) associated with its use. If these choices of thresholds go uncontested, wide uptake of novel treatments could prove detrimental to patients. In interpreting noninferiority thresholds, we will encourage you to use your own judgment rather than accepting that of the investigators, relieving you of the need to decipher what many may experience as obscure statistical reasoning used to define the thresholds.

Although others have explained the rationale and provided criteria for interpreting noninferiority trials,<sup>2,3,5-9</sup> this article strives to present a simple and practical approach based on Users' Guides principles. We will use contemporary examples to illustrate concepts that can guide optimal clinical practice. In doing so, we follow the 3-step approach of other Users' Guides, focusing on issues of validity, interpretation, and applicability of results specific to noninferiority trials (BOX 3).

**ARE THE RESULTS VALID?**

In previous Users' Guides, this question asks to what extent the results are likely to represent an unbiased estimate of effect vs systematic overestimates or underestimates. As with other studies addressing disease-management questions, noninferiority trials will reduce the risk of bias if they ensure concealed randomization; demonstrate balance of known prognostic factors; blind patients, clinicians, and outcome assessors; and ensure complete

**Box 1. Statistical Approaches to Setting an Acceptable Noninferiority Threshold**

There is no universally accepted method for defining an appropriate threshold. It very much depends on the eye of the beholder. Experts have recommended using sound statistical reasoning and clinical judgment in determining noninferiority thresholds.<sup>2,3</sup> What is sound reasoning for one observer, however, may strike another as misguided.

The US Food and Drug Administration (FDA) has produced draft guidance regarding noninferiority thresholds that has proved highly influential.<sup>4</sup> The logic of the FDA's approach begins by considering the smallest plausible benefit achieved by the existing standard treatment with which the novel—hopefully, noninferior—treatment is compared. One establishes the smallest plausible benefit of the existing standard treatment by examining the results of a trial of that treatment against the previous best care or placebo. To establish the smallest plausible benefit, one focuses on the CI around the observed estimate of effect (in technical terms, the CI around the point estimate) and, in particular, the boundary of the CI nearest to no effect.

For instance, the point estimate may suggest that the existing standard treatment decreases the absolute incidence of stroke, relative to placebo, by an absolute difference of 3%, with a 95% CI of 2% to 4% (FIGURE 2, top graph). The smallest plausible benefit of the standard drug is then 2%, or 2 fewer strokes for every 100 patients treated.

If the 95% CIs around the difference in strokes in a subsequent trial testing a novel drug for noninferiority include an increase in strokes of 2% (for instance, a point estimate of no difference, with a CI of a 2% decrease to a 2% increase), the results are consistent with the new drug being no better than placebo (Figure 2, scenario A). This is because the absolute benefit of the existing standard may be a reduction in strokes of as little as 2%, and those receiving the novel treatment may have a stroke rate of 2% more than the standard treatment—exactly equivalent to the rate on placebo.

The logic then goes that we should insist on some preservation of the treatment effect. Commonly, drug regulatory authorities stipulate that at least 50% of that minimal treatment effect be preserved. The threshold would, in this example, be 1%; if the novel treatment increases strokes by no more than 1% relative to the existing standard, at least 50% of the 2% absolute reduction in stroke has been preserved<sup>5,6</sup> (Figure 2, scenario B). Depending on the seriousness of the outcome, some may argue for retaining a greater proportion of benefit, resulting in a more challenging noninferiority margin. We have focused herein on expression of the noninferiority margin in absolute terms; sometimes, the choice of threshold is based on a relative rather than an absolute effect.

follow-up. Noninferiority trials are, however, vulnerable to misleading conclusions in ways that superiority trials are not. Although not strictly related to the risk of bias, we have classified the relevant concerns as issues of validity (Box 3).

**Did the Investigators Guard Against an Unwarranted Conclusion of Noninferiority?**

**Was the Effect of the Standard Treatment Preserved?** One way to achieve apparent noninferiority is to suboptimally administer the standard treatment. Suboptimal treatment can include enrolling patients less likely to be adherent or responsive to standard treatment; enrolling a population at low risk of the effectiveness outcome, particularly if the noninferiority threshold is expressed in absolute terms; reducing treatment intensity or administering treatment by a suboptimal route (eg, orally rather than intravenously); or terminating follow-up before treatment effects are fully manifest. One strategy to assess whether the treatment effect was

likely to have been preserved would be to evaluate the extent to which the design and conduct of the study attempted to overcome each of these threats to the standard treatment effect.

A second way to determine whether the effect of standard treatment has been preserved is to compare the event rate in the noninferiority trial with those seen in historical trials involving the standard treatment. A higher event rate in the standard treatment group in the noninferiority trial compared with the typical rate seen in historical trials would raise the suspicion of suboptimal administration of the standard treatment. Unfortunately, the competing explanation—prognostic differences between the populations enrolled in noninferiority vs historical trials—is also likely. Comparing patient characteristics between the trials could help decide which of the competing explanations is more likely, but the possibility remains that unmeasured prognostic features are responsible for the observed difference in event rates.

Take, for instance, the Rivaroxaban Once Daily Oral Direct Factor Xa Compared With Vitamin K Antagonism for Prevention of Stroke and Embolism in Atrial Fibrillation (ROCKET AF) trial, in which investigators declared rivaroxaban to be noninferior to warfarin in managing patients with atrial fibrillation.<sup>10</sup> Concerns exist about the extent to which the patients treated with warfarin remained within the therapeutic range of anticoagulation throughout this study in comparison with previous RCTs comparing warfarin with placebo. Investigators documented a mean time in therapeutic range (TTR) of 55% in the warfarin group in ROCKET AF, considerably less than rates of approximately 75% (range, 42%-83%) seen in prior studies<sup>11,12</sup> and in contemporary noninferiority trials.<sup>12</sup> Hence, we cannot be confident that the warfarin treatment effect was preserved in ROCKET AF. The apparent noninferiority of rivaroxaban to warfarin may be because the latter was suboptimally administered.<sup>13</sup>

**Figure 2.** Setting an Acceptable Noninferiority Threshold



A, Standard treatment decreases the absolute incidence of stroke, relative to placebo, by 3%, with a 95% CI of 2% to 4%. B, The blue dashed line represents the noninferiority zone. In scenario A, the 95% CIs around the difference in strokes between the novel treatment and the standard treatment includes an increase in incidence of strokes by as much as 2% with the novel treatment, thereby failing to retain 50% of the minimal treatment effect of the standard treatment. In scenario B, the same 95% CIs suggest that the novel treatment increases the incidence of strokes by no more than 1%, thus successfully preserving at least 50% of the 2% absolute reduction in stroke with the standard treatment.

Using the second criterion, the rate of stroke or systemic embolism in the warfarin group was lower in the ROCKET AF trial than has been seen historically,<sup>14</sup> despite the fact that patients in the ROCKET AF trial were older and had a higher prevalence of hypertension and diabetes mellitus than those in the previous trials.<sup>11</sup> Thus, event rates fail to support the suspicion of suboptimal warfarin administration in the control group. The low TTR, nevertheless, remains concerning.

**Did the Investigators Analyze Patients According to the Treatment They Received, as Well as to the Groups to Which They Were Assigned?**

A second issue has to do with how investigators dealt with patients who were randomized and followed up to the end of the study but did not take their medication as intended or did not use it at all. The purpose of randomization is to ensure that prognostic factors for the outcome of interest are balanced between treatment

groups. It is likely that those who do not adhere to the allocated treatment as set out in the study protocol are prognostically different from those who do.<sup>15</sup>

Investigators may be tempted to include only those individuals who were adherent to study protocol and omit those who were not (often called a per-protocol analysis). This is likely, however, to compromise the prognostic balance that randomization created in the first place. Since, more often than not, nonadherent patients are prognostically worse than adherent patients, the omission of those who failed to adhere to the novel treatment is likely to bias results toward an overestimation of treatment benefit in a superiority trial. In contrast, an analyze-as-randomized approach (intention-to-treat analysis) analyzes patients in the groups to which they were assigned irrespective of the level of patient adherence. As a result, it yields an unbiased—and typically more conservative—estimate of treatment effectiveness in a superiority trial.<sup>16</sup>

Unfortunately, the analyze-as-randomized approach has serious limitations in the context of noninferiority trials. Picture a noninferiority trial in which the novel treatment is actually substantially inferior to the current standard. Let us further suppose that, in this trial, many patients in the standard treatment group do not, for whatever reason, adhere to treatment. In the analyze-as-randomized approach, including these nonadherent patients may result in a substantial underestimate of the benefit of standard treatment and thus cause a misleading inference of noninferiority in comparison with the novel treatment.

The per-protocol analysis—which focuses only on those who use the treatment more or less as directed—while likely introducing prognostic imbalance can nevertheless provide some reassurance regarding noninferiority. If the results of such an analysis are consistent with those from the analyze-as-randomized approach and if both lie below the noninferiority threshold, our inference regarding noninferiority is strengthened. If, however, there are important differences between the results of the 2 analyses, the inference of noninferiority is weakened.

For example, the Cardiac Insufficiency Bisoprolol Study (CIBIS) III trial addressed the initial use of a  $\beta$ -blocker rather than an angiotensin-converting enzyme (ACE) inhibitor for preventing deaths or hospitalization in patients with heart failure.<sup>17</sup> The investigators set a noninferiority threshold of a 5% absolute increase in the primary end point of death or hospitalization with  $\beta$ -blocker use. The as-randomized analysis met their noninferiority threshold: the upper limit of the CI suggested that an increase in death or hospitalization greater than 4.4% with  $\beta$ -blockers was very unlikely. In the per-protocol analysis, however, the upper limit of the CI was 5.1%, just above the investigators' chosen threshold. Were one to accept the authors' threshold, the inference of noninferiority is weakened by the results of the per-protocol analysis. Whether one should accept the

authors' threshold at all is a point to which we will return.

### **Back to the Clinical Scenario: Validity of Inpatient vs Outpatient Pulmonary Embolus Trial**

In the pulmonary embolus study, the investigators randomized 344 patients with acute symptomatic pulmonary embolus at low risk of death to outpatient treatment for 5 or more days of inpatient treatment. Randomization was concealed via a central computer randomization system. Neither patients nor their caregivers were blinded to the allocated treatment, but adjudicators of outcome were. Patients in the treatment and control groups were similar with respect to known prognostic factors, including location of the embolus, comorbidity, and clinical findings. Complete follow-up was achieved in all but 5 patients. Although the lack of blinding raises concern, blinding of the outcome assessors provides a safeguard against risk of bias.

The crucial issues in optimal administration of the standard intervention in this study are the duration of time patients in the hospitalized group received LMWH and the TTR during subsequent warfarin treatment. Patients spent a mean of 8.9 days receiving LMWH, as long or longer than the standard in many settings (and thus satisfactory). The TTR was only 52%, which is suboptimal and raises concern. However, the TTR in the outpatient group was also 52%, substantially ameliorating the concern.

The investigators conducted both an analysis-as-randomized and a per-protocol analysis, which excluded patients in the hospitalized group discharged within 24 hours and those in the outpatient group discharged more than 24 hours after randomization. As you will see in the results that we present below, the per-protocol results do not substantially differ from the as-randomized results.

In conclusion, although the trial has some limitations in risk of bias, we would conclude moderate to high credibility of its findings.

### **WHAT ARE THE RESULTS?**

The relevant results of a noninferiority trial focus on: (1) the difference between novel and standard treatment in the effectiveness outcome that is the primary target of treatment, (2) the harm and burden outcomes that should favor the novel over the standard treatment, and (3) whether the results provide reassurance that the standard treatment was optimally administered.

### **Back to the Clinical Scenario: The Pulmonary Embolism Trial Results**

For pulmonary embolism, the primary effectiveness outcome is reducing recurrent VTE and the burden (staying in the hospital rather than being treated at home) is easily measured. Another important issue is the incidence of major bleeding, which could be conceptualized as an additional outcome warranting a noninferiority inquiry. Even if outpatient care was noninferior to inpatient care with respect to the primary effectiveness outcome, patients may choose to remain in the hospital if the risks of serious bleeding are substantially higher at home.

For each outcome, we are interested in the point estimate (the best estimate) of the difference in event rates between novel and standard treatments and its associated CI. The boundaries of the CI represent the range of plausible truth—less likely than the point estimate, but still plausible. Herein, we will focus on the absolute differences between groups at 90 days. In the as-randomized analysis, recurrent VTE occurred in 1 individual in the outpatient group and none in the inpatient group, a difference of 0.6% or 6 in 1000, with an upper boundary of the 95% CI of 2.7% (27 more VTE in 1000 outpatients). This result suggests that it is very unlikely that the recurrent VTE rate among outpatients is more than 4% (40 in 1000) greater than among inpatients ( $P=.01$ ), the authors' noninferiority threshold.

For serious bleeding, the investigators observed 3 events in the outpatient group and none in the inpatient

### **Box 2. Setting the Noninferiority Margin in the Pulmonary Embolism Trial**

Because there are no randomized trials comparing anticoagulation to no anticoagulation in pulmonary embolism, the authors could not use the procedure for setting the noninferiority margin described in Box 1. As an alternative, they first considered the likelihood of recurrent venous thromboembolism (VTE) at 90 days in low-risk inpatients with pulmonary embolism, which they estimated at 0.9%. They then specified a noninferiority margin of 4% (implying that patients would find it acceptable if the rate of recurrent VTE for outpatients would be less than 4.9%). They justified their choice by saying that it was similar to the noninferiority margins—3% to 5%—set in other trials of different anticoagulant regimens in acute VTE and outpatient vs inpatient treatment for deep venous thrombosis. The authors implicitly chose the same noninferiority margin—4%—for bleeding, though they provide no justification for this choice.

group (1.8%, or 18 in 1000 more bleeds in outpatients). The upper boundary of the 95% CI is 4.5%, which exceeds the authors' 4% threshold and therefore fails the statistical test of noninferiority ( $P=.09$ ).

The authors also present a per-protocol analysis. It is consistent with the as-randomized results, and the major bleeding outcome is actually more favorable to outpatient management (a difference of 1.2% favoring inpatient management, with an upper boundary of the 95% CI of 3.8%; the  $P$  value against the 4% threshold is .04).

### **How Can I Apply the Results to Patient Care?**

In applying findings from the medical literature to individual patient care, we suggest asking 3 questions (Box 3), of which one—assessing trade-offs be-

**Box 3. Users' Guides Approach to Evaluating a Noninferiority Trial****A. Are the Results Valid?**

Did novel and standard treatment groups start with the same prognosis?

Was prognostic balance maintained as the trial progressed?

Were the groups prognostically balanced at the completion of the trial?

Did the investigators guard against an unwarranted conclusion of noninferiority?<sup>a</sup>

Was the effect of the standard treatment preserved?

Did the investigators analyze patients according to the treatment they received, as well as to the groups to which they were assigned?

**B. What Are the Results?****C. How Can I Apply the Results to Patient Care?**

Were the study patients similar to my patient?

Were all patient-important outcomes considered?

Are the likely advantages of the novel treatment worth the potential harm and costs?<sup>a</sup>

<sup>a</sup>Includes issues specific to noninferiority trials.

are not always thorough in carrying out due diligence in this aspect (and others) of reporting.<sup>18</sup>

The risk that a trial reported as noninferiority may not have been planned as noninferiority again highlights the importance of an independent judgment of the noninferiority threshold. You may be tempted to turn to the authors of a study for guidance on assessing the key inferences from a noninferiority trial: are the advantages of the novel treatment worth the risks of loss of effectiveness? In doing so, you are implicitly accepting the authors' noninferiority threshold. For various reasons, investigators may have an incentive to be as lenient as possible with the choice of noninferiority threshold. Thus, accepting that threshold may not serve your patients' best interests.

Consider first the CIBIS-III trial investigating the substitution of  $\beta$ -blockers for ACE inhibitors in the initial treatment of heart failure that we used to illustrate the desirability of a per-protocol analysis. The as-randomized and per-protocol results straddled the authors' noninferiority margin of 5%. But is that margin appropriate? The harms or convenience advantages of  $\beta$ -blockers over ACE inhibitors are few, if any. Thus, patients are unlikely to accept starting with  $\beta$ -blockers if it really meant an absolute increase of up to 5% in the end point of death or hospitalization.

Consider next the Post-Operative Radiation Therapy for Endometrial Carcinoma 2 (PORTEC-2) trial that investigated the effect of vaginal brachytherapy (VBT) vs pelvic external beam radiotherapy (EBRT) on the primary outcome of vaginal recurrence of endometrial carcinoma.<sup>19</sup> The investigators set a noninferiority threshold of a risk difference of 6%—an increase in the primary outcome of 6 events in 100 patients—between the 2 groups at 5 years. After analyzing the data, they declared the VBT regimen to be noninferior to EBRT on the basis that the upper boundary of the CI—an absolute difference of 5%—fell below their

threshold. Although patients undergoing VBT report better health-related quality of life than those receiving EBRT,<sup>19</sup> for an outcome as serious as cancer recurrence, we suspect that few patients would be willing to choose the VBT approach if the actual increase was as great as 5%.

When looking at noninferiority trial results, we suggest you act as an advocate for your patients by assessing whether they are likely to consider the advantages of the novel treatment worth the potential loss in effectiveness. In doing so, you need to consider the confidence you have in the advantages of the novel treatment. Some advantages may be very clear (such as receiving treatment for pulmonary embolus at home rather than in the hospital). In other instances, advantages may be questionable or poorly documented.

However well documented the advantages of the novel therapy, trading off the desirable aspects with the potential loss in effectiveness may be a challenging judgment. It is not, however, fundamentally different from other patient-management decisions: they all involve trading off the desirable and undesirable consequences of the alternatives.<sup>20</sup> On the basis of your clinical experience and what you consider to be the values and preferences of most of your patients regarding the outcomes under consideration, try to estimate the noninferiority threshold: the maximum increase in risk of the primary outcome that your patients would, on average, be willing to accept in exchange for the novel treatment's reduction in harms or burden.

When you consider your patients as a group—ie, the typical patient—you may want to refer to published studies that provide insight into patients' values and preferences.<sup>21</sup> Having done this, now look at the upper boundary of the CI for the primary outcome and note the extent to which it exceeds your threshold. If the upper boundary is substantially greater than your threshold—as, in our judgment, it was in PORTEC-2—you may choose not to

tween a novel treatment's potential desirable and undesirable outcomes—includes issues specific to noninferiority trials.

**Are the Likely Advantages of the Novel Treatment Worth the Potential Harm and Costs?**

Is a particular noninferiority trial simply a failed superiority trial, portrayed to put a happy face on a sad result? When investigators plan their trials, they specify the analysis, and this specification has implications for how results are interpreted. It is the job of the editors to ensure that only trials planned as noninferiority are in fact reported as noninferiority trials. Unfortunately, editors

offer the novel treatment to your patients. However, if the upper boundary of the CI is near your threshold—that is, the balance between desirable and undesirable consequences is a close one—the decision must rest with each individual patient.

### Back to the Clinical Scenario: Applicability of the Pulmonary Embolism Trial Results

Your patient's clinical profile suggests a relatively low risk of death from pulmonary embolism. She would thus have been eligible for the trial, and its results are directly applicable to her care. Point estimates suggest similar and very low risks of recurrent VTE (6 in 1000); the difference in important bleeding is somewhat greater (18 more bleeds per 1000 in the outpatient group). The CIs raise more concern and include an increase in embolism of 2.7% (27 in 1000) and an increase in bleeding of 4.5% (45 in 1000), both within 90 days, in the outpatient care group.

Because their noninferiority margin for VTE has been met, the authors conclude that “[i]n selected low-risk patients with pulmonary embolism, outpatient care can safely and effectively be used in place of inpatient care.” Individuals who, all else being equal, would much prefer home treatment and are ready to focus on the point estimates that suggest that rates of adverse events (at least VTE) are likely similar with outpatient management, might agree. On the other hand, risk-averse individuals who perceive the possibility of increased risk of VTE and bleeding with outpatient management as not being worth the benefit of receiving treatment at home would not agree with this conclusion. We believe that there are likely to be a substantial number of such risk-averse individuals. Reliance on the authors' noninferiority would not serve such patients well.

### CONCLUSIONS

Critical appraisal of noninferiority studies closely follows the principles and

criteria for assessing any study of novel management strategies. With respect to validity, they require special attention to the optimal use of the standard treatment and to the results of the as-randomized and per-protocol analyses. With respect to the trade-offs between desirable and undesirable consequences, they require close attention to best estimates and CIs around the difference in effectiveness outcomes between novel and standard treatments. In particular, clinicians should consider whether patients would be willing to accept loss in the effectiveness outcome suggested by the upper boundary of the 95% CI, irrespective of whether this interval lies below or above the investigators' choice of noninferiority threshold.

**Author Contributions:** Dr Guyatt had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Financial Disclosures:** All authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none were reported.

### REFERENCES

- Aujesky D, Roy PM, Verschuren F, et al. Outpatient versus inpatient treatment for patients with acute pulmonary embolism: an international, open-label, randomised, non-inferiority trial. *Lancet*. 2011;378(9785):41-48.
- Fleming TR. Current issues in non-inferiority trials. *Stat Med*. 2008;27(3):317-332.
- Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med*. 2006;145(1):62-69.
- Temple R, O'Neill R. *Guidance for Industry Non-Inferiority Clinical Trials*. Rockville, MD: Food and Drug Administration, Department of Health and Human Services; 2010.
- Le Henaff A, Giraudeau B, Baron G, Ravaud P. Quality of reporting of noninferiority and equivalence randomized trials. *JAMA*. 2006;295(10):1147-1151.
- Piaggio G, Elbourne DR, Pocock SJ, Evans SJW, Altman DG; CONSORT Group. Reporting of noninferiority and equivalence randomized trials: extension of the CONSORT 2010 statement. *JAMA*. 2012;308(24):2594-2604.
- Scott IA. Non-inferiority trials: determining whether alternative treatments are good enough. *Med J Aust*. 2009;190(6):326-330.
- Götzsche PC. Lessons from and cautions about non-inferiority and equivalence randomized trials. *JAMA*. 2006;295(10):1172-1174.
- Schumi J, Wittes JT. Through the looking glass: understanding noninferiority. *Trials*. 2011;12:106.
- Patel MR, Mahaffey KW, Garg J, et al; ROCKET AF Investigators. Rivaroxaban versus warfarin in non-valvular atrial fibrillation. *N Engl J Med*. 2011;365(10):883-891.
- Jackson K, Gersh BJ, Stockbridge N, et al; Duke Clinical Research Institute/American Heart Journal Expert Meeting on Antithrombotic Drug Development for Atrial Fibrillation. Antithrombotic drug development for atrial fibrillation: proceedings, Washington, DC, July 25-27, 2005. *Am Heart J*. 2008;155(5):829-840.
- Granger CB, Alexander JH, McMurray JJ, et al; ARISTOTLE Committees and Investigators. Apixaban versus warfarin in patients with atrial fibrillation. *N Engl J Med*. 2011;365(11):981-992.
- Fleming TR, Emerson SS. Evaluating rivaroxaban for nonvalvular atrial fibrillation—regulatory considerations. *N Engl J Med*. 2011;365(17):1557-1559.
- Hart RG, Benavente O, McBride R, Pearce LA. Antithrombotic therapy to prevent stroke in patients with atrial fibrillation: a meta-analysis. *Ann Intern Med*. 1999;131(7):492-501.
- Kunz R, Guyatt G. Which patients to include in the analysis? *Transfusion*. 2006;46(6):881-884.
- Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ*. 2001;165(10):1339-1341.
- Willenheimer R, van Veldhuisen DJ, Silke B, et al; CIBIS III Investigators. Effect on survival and hospitalization of initiating treatment for chronic heart failure with bisoprolol followed by enalapril, as compared with the opposite sequence: results of the randomized Cardiac Insufficiency Bisoprolol Study (CIBIS) III. *Circulation*. 2005;112(16):2426-2435.
- Yank V, Rennie D, Bero LA. Financial ties and concordance between results and conclusions in meta-analyses: retrospective cohort study. *BMJ*. 2007;335(7631):1202-1205.
- Nout RA, Smit VT, Putter H, et al; PORTEC Study Group. Vaginal brachytherapy versus pelvic external beam radiotherapy for patients with endometrial cancer of high-intermediate risk (PORTEC-2): an open-label, non-inferiority, randomised trial. *Lancet*. 2010;375(9717):816-823.
- Guyatt GH, Haynes RB, Jaeschke RZ, et al; Evidence-Based Medicine Working Group. Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. *JAMA*. 2000;284(10):1290-1296.
- MacLean S, Mulla S, Akl EA, et al; American College of Chest Physicians. Patient values and preferences in decision making for antithrombotic therapy: a systematic review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2012;141(2)(suppl):e1S-e23S.