Planned Equivalence or Noninferiority Trials Versus Unplanned Noninferiority Claims: Are They Equal?

Benny Chung-Ying Zee, Centre for Clinical Trials, School of Public Health, Chinese University of Hong Kong, Shatin, Hong Kong

There are many reasons why equivalence or noninferiority trials are being planned. Clinical trials in cancer often have several measures of outcome, such as survival, progression, treatment toxicity, and quality of life. It is rare that the same treatment is superior for all outcomes. Therefore, benefits and risks need to be balanced against one another. A small loss in survival may be acceptable for a large gain in the reduction of toxic effects. Therefore, the objective of equivalence trials is to demonstrate that a new treatment is equivalent to a standard therapy with regard to a specific clinical end point and has an intrinsic benefit for other clinical end points. A noninferiority trial refers to a study in which the primary objective is to evaluate whether the new treatment is not inferior to or as effective as the standard therapy for a particular end point. The new treatment may be less invasive and less debilitating, or it may be less expensive. Once noninferiority with respect to the primary end point has been demonstrated, it would be an attractive treatment option to patients and to the health care system in general.

A planned equivalence or noninferiority trial requires a properly formulated question and an assessment procedure. There are, in general, two approaches for the assessment. They properly formulated question and an assessment procedure. Therefore, the objective of equivalence trials is to demonstrate that a new treatment is equivalent to a standard therapy with regard to a specific clinical end point and has an intrinsic benefit for other clinical end points. A noninferiority trial refers to a study in which the primary objective is to evaluate whether the new treatment is not inferior to or as effective as the standard therapy for a particular end point. The new treatment may be less invasive and less debilitating, or it may be less expensive. Once noninferiority with respect to the primary end point has been demonstrated, it would be an attractive treatment option to patients and to the health care system in general.

A planned equivalence or noninferiority trial requires a properly formulated question and an assessment procedure. There are, in general, two approaches for the assessment. They are hypothesis testing and CI approaches. Both approaches require a prespecified quantity $\delta$, which is called an equivalence limit in an equivalence trial and a noninferiority margin in a noninferiority trial. The noninferiority margin reflects the degree of inferiority of the new treatment compared with the standard treatment that the trials attempt to exclude. The choice of $\delta$ is critical, in that it affects the sample size calculation and the conclusion of the study. The noninferiority margin $\delta$ should be chosen as the largest clinical difference or degree of inferiority that one would be willing to accept with the use of the new treatment. A general rule of thumb is that this quantity must be smaller than the minimal clinical differences usually used to calculate sample size in a superiority trial. Hence, the sample size required in a noninferiority trial is larger than that of a superiority trial. Because most clinicians find the CI approach easier to understand, we illustrate the concept of equivalence and noninferiority using the CI approach. Figure 1 shows the estimated log hazard ratio $[\ln(\theta)]$ between the new treatment and the standard treatment, that is, $\ln(\theta) = \ln(\lambda_1) - \ln(\lambda_2)$, where $\lambda_1$ and $\lambda_2$ are hazard rates for the new treatment and the standard treatment, respectively. When the lower and upper bounds of the 95% CI for the log hazard ratio lies completely within the equivalence limits (ie, between $-\delta$ and $\delta$), we would conclude that the two treatments are statistically equivalent. For a noninferiority trial, the upper 90% CI for the estimated log hazard ratio would be used. We would conclude that the new treatment is not inferior to the standard treatment if the upper 90% confidence limit is less than the noninferiority margin $\delta$.

Two examples have been chosen to illustrate the planned noninferiority trials; these trials were chosen because they have different results leading to different conclusions, but both contained some important common characteristics shared by most properly planned noninferiority trials. The first study was a randomized trial by Ozols et al comparing cisplatin and paclitaxel versus carboplatin and paclitaxel in patients with optimally resected stage III ovarian cancer. A total of 792 patients were randomly assigned to receive either cisplatin 75 mg/m$^2$ plus a 24-hour infusion of paclitaxel 135 mg/m$^2$ or carboplatin (area under the curve $= 7.5$) intravenously plus paclitaxel 175 mg/m$^2$ over 3 hours. The noninferiority margin was selected as a hazard ratio of 1.25. The hazard ratio for time to progression was 0.88 (95% CI, 0.75 to 1.03), and the hazard ratio for time to death was 0.84 (95% CI, 0.70 to 1.02). More than 85% of patients received six cycles of treatment. Because the upper bounds of the 95% CIs for both end points were less than 1.25, the study concluded that carboplatin plus a 3-hour infusion of paclitaxel was not inferior when compared with cisplatin plus a 24-hour infusion of paclitaxel. In addition, there was less toxicity in the carboplatin arm, and the 3-hour infusion of paclitaxel was easier to administer.

The second example was a randomized trial by Lukka et al comparing two fractionation schedules for patients with localized prostate cancer. A total of 936 early-stage prostate cancer patients were randomly assigned. This study was designed to determine whether a hypofractionated treatment regimen of 52.5 Gy in 20 fractions administered over 28 days (short arm) was as effective as the conventional treatment regimen of 66 Gy in 33 fractions administered over 45 days (long arm). The primary end point was the rate of biochemical or clinical failure, with a noninferiority margin of $-7.5\%$. The 5-year biochemical or clinical failure probability was 52.95% and 59.95% in the long and short arms, respectively (a difference of $-7.0\%$; 90% CI, $-12.6\%$ to $-1.4\%$). Approximately 2% of patients did not receive the prescribed radiation or received slightly altered fractionation, but all patients received...
complete courses of radiation treatment. The final conclusion was that the hypofractionated radiation regimen was indeed inferior to the standard regimen. Although the two examples have different outcomes (one showing noninferiority and the other showing inferior results), both have a lot in common in terms of their outcomes (one showing noninferiority and the other showing the standard regimen). Although the two examples have different outcomes, the hypofractionated radiation regimen was indeed inferior to the standard regimen. The conclusion was that the hypofractionated radiation regimen was indeed inferior to the standard regimen.

Both studies have properly calculated sample size using methods based on the noninferiority design. The compliance of both studies was reported, and both of the studies have relatively high treatment compliance rates.

In this issue, Steensma et al reported a randomized trial undertaken to determine whether epoetin alfa 120,000 U administered every 3 weeks was inferior to standard weekly maintenance epoetin alfa treatment at a dose of 40,000 U in patients with cancer-associated anemia. All patients received weekly induction of epoetin alfa regimen for the first 3 weeks. The primary study end point was the proportion of patients receiving RBC transfusion. Transfusions were administered only for hemoglobin \( \leq 8.0 \text{ g/dL} \), unless patients were judged by the treating clinician to have intolerable or dangerous ischemic symptoms at a higher level of hemoglobin. Secondary end points included changes in hemoglobin and quality of life from baseline measurements. The conclusion of this study was that epoetin alfa can be safely administered every 3 weeks at an equivalent total dose without increasing transfusion needs or sacrificing quality of life. In other words, the 3-week treatment schedule is not inferior compared with the standard weekly regimen. The noninferiority conclusion was based on the transfusion rate of 5% difference in favor of the new treatment arm, with a 95% CI of \(-4\%\) to \(13\%\). Although the study was designed as a superiority trial to detect a 13% difference in transfusion between the two arms, a noninferiority conclusion was drawn after knowing that there was no statistically significant difference between the two arms. Because the original design was asking for less than 10% difference in transfusion, and the difference was not statistically significant, the study was stopped before the end of the planned duration. The conclusion was that epoetin alfa can be administered every 3 weeks without increasing transfusion needs or sacrificing quality of life.

Superiority trial versus noninferiority trial. In conclusion, it would be appropriate to use an as-treared analysis instead of ITT analysis. Interim analysis for equivalence and noninferiority trials is also a topic requiring special attention. In a recent article, Korn et al discussed the appropriate time to release data for data monitoring in a noninferiority trial. Implementation and the rationale of data monitoring in a noninferiority trial is clearly different from that of a superiority trial design. A formal data and safety monitoring committee should seek clarification during the data-monitoring process to properly classify the study as having a superiority design or having a noninferiority design. The frequency of monitoring and the advice from the data and safety monitoring committee would then be able to tailor to the specific needs of the study accordingly.

Greene et al examined publications from 1992 to 1996 through a MEDLINE search and found that 67% of published articles claimed equivalence after a failed test for comparative superiority, and in 10% of the studies, the claim of equivalence was not statistically evaluated. The sample size from which to confirm equivalence results had been calculated in advance for only 33% of the reports. The problem of drawing unplanned equivalence or noninferiority conclusion is subtle, but the implication may be large in clinical applications based on the interpretation of these results. In conclusion, it would be
difficult to consider an unplanned noninferiority conclusion as having the same level of evidence compared with the evidence of a properly planned noninferiority trial.

© 2006 by American Society of Clinical Oncology

REFERENCES

Author’s Disclosures of Potential Conflicts of Interest
The author indicated no potential conflicts of interest.

Author Contributions
Conception and design: Benny Chung-Ying Zee
Manuscript writing: Benny Chung-Ying Zee
Final approval of manuscript: Benny Chung-Ying Zee