Trend Toward Noninferiority Trials May Mean More Difficult Interpretation of Trial Results

By Rabiya S. Tuma

The number of noninferiority trials in oncology is on the rise, and the trend will probably continue as the number of approved drugs increases. Experts warn that this shift from testing superiority to noninferiority will make interpreting trial results more difficult. Some researchers even question the value of dedicating limited resources just to show that a new drug is no worse than one already in use.

Most oncology trials remain superiority trials, which are designed to establish that a new treatment is outright better than the existing therapy. The increased use of noninferiority trials, which are designed to establish that the new therapy is no worse than the existing one, means that researchers and clinicians need to become more familiar with the approach, one that experts warn is not always intuitive—or properly used.

There is a perception among some clinicians and researchers that pharmaceutical companies use noninferiority trials as a way to get regulatory approval for a drug that offers nothing better than what is already on the market. But there are legitimate scientific reasons for some noninferiority trials, according to Edith Perez, M.D., director of the Mayo Clinic’s multidisciplinary breast clinic in Jacksonville, Fla.

“We are looking for better therapies for our patients,” Perez said. “That needs to be balanced with the possibility of finding therapies that are equally efficacious in terms of disease-free survival but that have other potential benefits that are more, if I may say, intangible. That is where noninferiority trials may play a role.”

Perez is currently leading a large phase III trial, called ALTTO (Adjuvant Lapatinib and/or Trastuzumab Treatment Optimization), which is designed to compare lapatinib alone, trastuzumab alone, a combination of the drugs, and the two drugs in tandem in women whose tumors overexpress Her2 and who have recently completed chemotherapy. The primary goal of the trial is to find a regimen that is better than trastuzumab alone, the current standard of care. However, the researchers also designed the trial to include a noninferiority comparison between the two single-agent arms. So even if none of the arms improves disease-free survival relative to trastuzumab, but the lapatinib arm is no worse, the trial could still be a positive trial—and grounds for regulatory approval of lapatinib as adjuvant therapy.

“If these drugs are found to be equivalent, for some people that would be a positive result because lapatinib is not intravenous,” Perez explained. “And it has some potential reduced costs, in terms of treatment time and in terms of the drug itself.”

Choosing the Right Design

A failed superiority trial is not enough to declare a drug noninferior. Noninferiority trials have particular statistical requirements. This fact appears to be lost on many researchers, as well as some journal editors and reviewers. In one of the few studies that evaluated the quality of noninferiority trials, Yale University researchers found that only 51% of the reports published between 1992 and 1996 that claimed equivalence were designed to answer the question. Moreover, authors of 67% of the examined reports simply declared that the two trial arms were equivalent after a standard test for superiority showed a nonsignificant difference. The problem with that approach, the Yale authors point out, is that an infinite number of patients would be needed to statistically show that two treatments are equivalent. A superiority trial design is dramatically underpowered to answer a question about noninferiority.

Trials can be designed to test both superiority and noninferiority, like the ALTTO trial, but researchers must accommodate both goals in the initial trial design, said Boris Freidlin, Ph.D., a researcher in the biometric research branch of the Division of Cancer Treatment and Diagnosis at the National Cancer Institute. These hybrid designs, as he and his colleagues refer to them in the Journal of Clinical Oncology, require more patients than a standard superiority design. And such designs are really appropriate only when researchers expect only a marginal improvement in clinical outcome with the new agent.

Getting to the Statistics

The biggest problem is that nonexperts don’t really know what noninferiority trials mean. “These trials are hard to understand, hard to explain, and if you are not used to them, they are hard to read,” said Anita Das, Ph.D., senior director of biostatistics at...
Cerexa Inc., in Alameda, Calif. “You have to understand these trial designs when you read the data. If you don’t, and you are used to superiority trials and you read the data as you have been taught to for drug A is better than drug B, you may misinterpret it because your thinking is kind of backward on this.”

An example of the backward thinking required when interpreting these trials is the role of patient compliance. In a superiority trial, poor compliance will dilute any benefit of the new regimen, making it less likely that the trial will be successful. But in noninferiority trials, poor compliance has the opposite effect: It reduces any difference between the trial arms, making the treatments appear more similar than they might actually be.

Because of that issue, reports on well-executed noninferiority trials should state clearly the compliance rate in each arm, Das said. She also thinks that authors should perform both an intention-to-treat analysis, a standard in superiority trials, and an analysis including only those patients who received the treatments as designed.

But compliance isn’t the only drawback, Freidlin said. “Noninferiority trials are more complex from a statistical standpoint, as well as from a clinical evaluation standpoint.” When evaluating a superiority trial, researchers compare the median estimated effect—the point estimate—and test whether there is a statistically significant difference between the two arms. By contrast, in a noninferiority trial, the median effect in the control arm is set equal to zero, and the experimental arm is plotted relative to that. For example, if the median effect of the experimental arm is 2% less than that obtained in the standard arm, it would be plotted as −2%.

Comparing the point estimate of the benefit in a noninferiority trial is only one aspect of the equation. The more important part of the data is the 95% confidence interval, both ends of which must fall within a specified boundary, called the noninferiority margin, for the trial to be a success.

Importantly, the noninferiority margin must be determined separately for each trial and depends on the amount of benefit gained from the standard therapy over placebo. For example, in the INTEREST (Iressa Non–Small-Cell Lung Cancer Trial Evaluating Response and Survival Against Taxotere) trial, researchers wanted to show that gefitinib (Iressa) was not inferior to docetaxel in previously treated non–small-cell lung cancer patients. In a previous trial, TAX-317, docetaxel increased overall survival by 2 months over best supportive care. On the basis of that finding, the INTEREST researchers decided that they would be willing to lose up to 50% (1 month) of that benefit to gain a drug with fewer side effects. The trial met that goal.

The difficulty is that the estimated benefit for the standard treatment is often not based on a placebo-controlled trial, Freidlin said. Then, setting the noninferiority margin is difficult. “That is the part of the design that the majority of studies are missing,” Freidlin said. “There are a lot of studies that claim noninferiority and just give a confidence interval from the trial, saying, ‘We can’t rule out a 20% increase in mortality.’ But they don’t have a good presentation of what the benchmark is, what the benefit of the standard treatment is that you are trying to preserve.”

Without that benchmark, researchers can end up with a proposed therapy that is worse than no treatment at all. For example, if the benefit of the standard treatment is a 15% improvement over placebo and the new treatment is 20% worse than the new one, it would be unacceptable “because it means you are potentially doing worse than no treatment,” Freidlin said.

Both Freidlin and Das point out that a common problem in noninferiority trials is the number of patients required. Because the acceptable difference between the two arms is typically smaller by half than what was gained moving from a placebo to a standard treatment, the number of patients required to maintain the same statistical power increases substantially. On average, fourfold more patients are required for a noninferiority design than for a superiority one, according to Freidlin.

Is Noninferiority the Question?

Whether noninferiority should be the question is one objection that clinicians have to noninferiority trials. “We have hundreds of new drugs out there, and the patient population who can go on trial is not limitless,” said Kathy Albain, M.D., professor of medicine and oncology at the Loyola University Chicago Stritch School of Medicine. With more approved drugs available, either companies now need to test the superiority of an agent in patients who have already failed many therapies—a setting in which even a good drug may fail because the disease is so resistant to therapy—or they can run a noninferiority trial to show that the drug works just as well as standard therapies in less heavily pretreated patients. Unless companies have strong evidence that a new agent is substantially better than the existing agent, they won’t want to risk a failure in a superiority trial.

“That creates a problem when we have so many exciting new agents that we want to get onto the menu a bit sooner than these very cumbersome noninferiority trials allow. In some certain circumstances those [trials] are necessary, but what many of us are hoping for are some novel creative designs that will be accepted by all parties—clinical researchers, industry, and the [U.S. Food and Drug Administration]—that will allow us to move the new agents into use as quickly as possible, as quickly as we establish safety and efficacy, without having to do several thousand patient trials for each of them.”

Until better designs are available, however, several groups are working to improve their quality. The FDA is working on guidelines about noninferiority trials, but they will not be available until the fall of 2008. The agency declined to discuss any details. The European Medicines Agency has already developed guidelines for such trials, which it posted in January 2006.

Also, the CONSORT (Consolidated Standards of Reporting Trials) group has modified its original guidelines, which were developed for superiority trials, to cover noninferiority trials (http://www.consort-statement.org/index.aspx?o=1049). Although the group does not yet have direct evidence that the design and execution of noninferiority
NEWS

trials has improved since they published the revised guidelines last year, one group member is hopeful. “I sense that there is an underlying trend toward better planning and reporting of noninferiority trials,” said Stuart Pocock, Ph.D., professor of medical statistics at the London School of Hygiene and Tropical Medicine. “Trialists are better aware; FDA and other regulators request appropriate designs; and articles, such as ours, no doubt help.”

For now, researchers seem to accept that noninferiority trials have a crucial, if limited, role in oncology drug development. “We don’t often do noninferiority trials in the setting of a National Cancer Institute–sponsored trial, because we are looking for something that is better,” said David Gandara, M.D., director of the thoracic oncology program and associate director for clinical research at the University of California Davis Cancer Center in Sacramento. “But [the design] is not inappropriate at all. Statistics are one way to help answer a clinical question; noninferiority designs concentrate on one aspect of a comparison when you expect that other outcomes will differ.” If a new therapy has less or different toxicity or is easier to administer, then looking for ‘no worse than’ can be an important outcome for patients, Gandara and others said.

The catch, though, is that the trial design needs to be used appropriately, both for methodology and clinical setting—and the audience must be sure that those criteria are met before accepting the results. Overall, Freidlin said, this is a case where readers should take a bit of extra time to decide just exactly what the data mean.

© Oxford University Press 2007. DOI: 10.1093/jnci/djm258