Large Trials, Small Gains: Is Change on the Way?

By Rabiya S. Tuma

arge phase III trials designed to detect small differences in progression-free or overall survival have become common in oncology. Yet a growing number of oncologists are critical of the approach and think the community needs to aim for substantially larger gains to make meaningful strides in patient care.

No one expects that large trials powered to detect statistically significant small gains will disappear, but academic and industry experts think that pressure to look for bigger gains will increase. Proposed fixes include raising the standards for drug approvals, putting more emphasis on clinically meaningful gains, and designing trials to find big gains in small, biomarker-defined subgroups.

Antonio Tito Fojo, M.D., Ph.D., head of the experimental therapeutics section at the National Cancer Institute, realized 2 years ago that a problem existed when he heard the results of a trial that compared chemotherapy alone to chemotherapy plus cetuximab in patients with advanced non-small cell lung cancer. With 1,125 patients enrolled, the randomized trial demonstrated a 1.2-month increase in overall survival, from a median of 10.1 months in the chemotherapy-alone arm to 11.3 months in the experimental arm. The difference was statistically significant, and thus the research community considered the trial a success.

Raising the Bar

Fojo questions that interpretation, suggesting that the bar that measured success in this trial and others like it has been lowered too far. “We say the median may seem low, but some patients got more benefit, so the low bar is okay,” he said. “But I think, in that case, one has to begin to be concerned that some patients are not only not being helped but may actually be harmed.”

Fojo pointed out that if the median survival is 1.2 months, then many patients are at the lower end of the spectrum. And that lower end extends not just to zero days gained but also into days lost. He pointed out that most researchers think in terms of zero days, or up, but the curve extends below that point, which is often forgotten.

When asked why the research community has lowered the bar, Fojo and others said that they think the problem starts with the shift in objectives—from curing cancer to prolonging life. “The moment you set that as the goal—prolonging life instead of a cure—then you get to the point of asking what is a meaningful prolongation of life,” Fojo said. Most would agree that 1 year or 6 months is meaningful. But when the median difference is much less than that, the value of the treatment is less certain.

At this point, researchers often argue that some patients—those on the higher end of the spectrum—had more substantial benefit and therefore the treatment is worth trying. Fojo counters by pointing to the lower end of the scale: “As the bar gets lower and lower, you run the risk of a greater number of patients being harmed,” he said. “We have gotten to think that what we do doesn’t harm, but it does.”

Laurence Baker, D.O., professor of internal medicine at the University of Michigan Medical School and chair of the Southwest Oncology Group (SWOG), says that the bar is not high enough. He argues that a system detects only what is designed to detect. “Cure is clearly the expectation of society,” he said. “My criticism is that we are not taking that seriously enough. We have a system that doesn’t even really try to meet that expectation.”

But for clinical trial groups, whose primary goal is improving patient care, that shouldn’t be acceptable. “We should say, ‘Our goal is to improve the care of cancer patients in a clinically meaningful manner,’” Baker said.

In reviewing past SWOG trials, Baker said he was disturbed to find few that really changed the practice of medicine. For future trials, he is asking SWOG researchers to at least ask themselves how their proposed trial could alter practice. And for trials designed to evaluate the efficacy of a new therapy or therapeutic approach, he wants SWOG researchers to poll their expert colleagues to come up with some sort of consensus about what would be a meaningful gain in a given disease setting—and then design the trial accordingly.

William R. Sellers, M.D., vice president and global head of oncology at the Novartis Institutes for Biomedical Research in Cambridge, Mass., agrees that the community needs to carefully consider how much time gained is meaningful. And although relatively small incremental gains have sometimes added up to substantial improvements, in adjuvant breast and colorectal cancer for example, “the failure rate of that approach has been mounting over time,” Sellers said.

That approach could even get in the way of finding a cure, according to Sellers. “One of my concerns is that we will build up complicated regimens that actually aren’t going to be the foundation for a cure, and at the same time you are making it harder to come in with a curative regimen,” he said. For example, a complicated regimen may be so close to the maximum toxicity that patients can handle that researchers cannot add another agent, and persuading researchers to drop an established agent from the regimen to make room for a novel one can be hard.

Statistical Issues

Of course, no one sets out to find a 3- or 4-week increase in survival. The trials are designed with a longer benefit in mind, and they include enough patients so that researchers can be sure that they will see the desired change if it occurs. As a result, the trial ends up with
enough power to report much smaller gains as statistically significant, according to both academic and industry statisticians.

That is part of what happened in a phase III trial that led to the approval of erlotinib for treating advanced pancreatic cancer, according to James Reimann, Ph.D., director of oncology biostatistics at Genentech in South San Francisco, Calif. The trial compared gemcitabine to gemcitabine plus erlotinib, and researchers designed it to find a median increase in overall survival of 1.8 months, which pancreatic cancer experts viewed as clinically significant. When the results appeared, however, the 569-patient trial showed a much smaller, but still statistically significant, increase in overall survival. The difference was just 0.33 months, with a median of 6.24 months in the erlotinib-plus-gemcitabine arm and 5.91 months in the gemcitabine-alone arm.

“I think there are two things going on,” Reimann said. “The observed effect was a bit weaker than what was designed for, and the median doesn’t look that profound. But when you look at the curves as a whole, it looks stronger.” He noted that the hazard ratio (a measure of how often over time a particular event, such as death, happens in one group compared to how often it happens in another group) is a more reliable assessment of benefit. Here it was 0.82 for risk of death with the addition of erlotinib.

Richard Pazdur, M.D., director of the office of oncology drug products at the U.S. Food and Drug Administration, agreed that the hazard ratio is a better measure of benefit in this instance. And, he noted, the agency generally looks more favorably on improvements in overall survival than time to progression or progression-free survival, so that even small gains in overall survival can be enough to support a drug’s approval.

Regarding small differences in progression-free survival, Pazdur pointed out that one reason that some trials detect such differences is that the agency wants to see both overall and progression-free survival. That means that even if the primary endpoint of a trial is progression-free survival, the researchers must enroll enough patients for an overall survival analysis. And that means the trial is essentially

continued on page 1223
overpowered for its primary endpoint and thus can detect very small changes that may or may not be clinically meaningful.

Pazdur emphasized, however, that just showing statistical significance is not sufficient. “We are not a slave to a P value,” he said. “There is a difference between clinical significance and statistical significance, and we should not equate them. It has to be a clinically valuable difference.”

Bigger Gains, Smaller Groups?

One benefit of large trials is that their sample size lets researchers look for subgroups that derived benefit. The problem, according to several researchers, is that such analyses are rarely done. “In very few instances did anyone go back and try to figure out if, within this unselected group that had this small benefit, there was a smaller group of patients that had all the benefit,” Sellers said. In fact, that has been relatively difficult to do until recently because the community has not been rigorous about collecting tissue samples and the technical tools for the analysis have been lacking. “It is an evolution,” Sellers said. “A lot of those technical hurdles are changing. I think the speed with which we could do subgroups analysis could be much faster in the future.”

Sellers thinks drug developers and researchers should be looking for big gains, which even small trials can detect. For example, if half the patients with a given biomarker have a substantial reduction in their tumor burden in response to a new treatment, a phase II trial would easily see that outcome. And, he added, such regimens should be used as building blocks for curative regimens for that select group of patients.

Sellers thinks the community is going to be pushed in that direction for a variety of reasons. For pharmaceutical companies, economic pressures are going to drive the change. As the cost of care increases, insurance companies and other payers will demand bigger benefits and be willing to pay for them, but they will not be willing to pay for small, marginal gains. As for the cooperative groups and other academic researchers, he thinks the change is likely to come from the inside. “They could shift the field themselves by saying, ‘We are not doing 3,000-person randomized trials. We’re going to look for big effects with smaller numbers of patients.’”

SWOG’s Baker is trying to push the cancer research community that way, or at least provoke thought. “I am trying to get people to stop saying how successful the cancer research enterprise is. It is not true. It is just not true,” he said. “Look at the kinds of resources being spent and ask yourself if they are well spent or if they could be spent better. And if you want to improve the system, the first thing you have to do is examine where it is weak; you have to be self-critical.”

© Oxford University Press 2010. DOI: 10.1093/jnci/djq322