Trends-in-Medicine

September 2002
By Lynne Peterson

SUMMARY

Uncertainty over the best endpoints for a clinical trial is not limited to researchers. Even the FDA has been holding internal meetings to discuss appropriate endpoints, particularly in oncology trials. The FDA has approved drugs based on trials with a time-to-progression (TTP) endpoint, but it is a high hurdle, and the agency has no immediate plans to issue guidelines on the use of TTP endpoints. TTP is less precise than survival, and it raises questions -- about patient selection, cherry-picking, bias, measurement interpretation, the comparator, etc. – that must be answered.

Trends-in-Medicine
Stephen Snyder, Publisher
1879 Avenida Dracaena
Jensen Beach, FL 34957
772-334-7409 Fax 772-334-0856
www.trends-in-medicine.com

Using Time-to-Progession as a Clinical Trial Endpoint

The choice of endpoints for a clinical trial is a critical decision, and it can affect the approvability of a drug or device by the FDA. Choosing the wrong endpoints can mean a company will have to do an additional study (or more than one study) before the FDA will approve its product. To answer questions that have been raised about the use of time-to-progression (TTP) as a primary endpoint, particularly in oncology trials, six oncologists, a senior FDA official and other FDA sources were interviewed.

Companies too often make the mistake of assuming that FDA approval of a trial protocol means that a positive trial will assure approval. An FDA official said, "That is a common source of misunderstanding. When we say a protocol may proceed, it is no guarantee that if that protocol is done, that it will lead to approval. We evaluate a particular protocol as an ethical approach for research. So, we look at it as permitting a protocol to proceed rather than as approval of a protocol. I've actually had companies have to retract advertisements because occasionally they say they are doing an 'FDA-approved study,' which sounds like the drug is approved when we just mean we allowed the study to proceed."

Choosing the wrong endpoints can be a painful and expensive mistake. It's a lesson that Medtronic and Guidant learned. Both of those companies got permission to start clinical trials of their biventricular pacing devices for congestive heart failure. Those trials did not include a mortality component, the FDA approved the trial designs, and officials of both companies insisted that meant the FDA was not requiring mortality data for approval. However, FDA officials, speaking at medical meetings, made it quite clear that mortality was a big concern for this new technology, especially after the experience with milrinone (Sanofi's Primacor), a congestive heart failure drug which was approved because it made patients feel better but which later was found to decrease survival. However, eventually, Guidant had to stop its trial, enroll more patients, and collect mortality data. Medtronic actually had to complete an additional trial. It was an expensive lesson for both companies.

There is a way for companies to be sure they are designing a trial with a good chance of approval. The FDA has a program called Special Protocol Assessments. In that program the FDA makes a commitment to honor that design as the basis for registration – but not commit to honor the results unless they are interpretable and positive. An official said, "The whole rationale for Special Protocol Assessments is to have agreement of everyone involved on how to minimize bias and uncertainty in the trial and to have agreement on the endpoints and the planned analysis. (In the past, we have seen some companies) that will do a study, analyze the bejeebers out of it, and come out with something somewhere to show a difference. That's not fair to patients, to us or to the public."

There have been internal discussions at the FDA over the past couple of years about endpoints. A group of oncology medical officers from both drugs and biologics have been meeting every other week to discuss endpoints in different malignancies. The FDA official said, "What we would like to do is have that discussion take place with other partners and in other ways, bringing in other people over time, and not just in one venue. We don't want to pick one organization with which to partner to have that be the official endpoint discussion. Rather, we want to systematically look at endpoints, so we are better informed and can discuss it with many people in many settings. We recognize that endpoints can be dynamic and fluid."

With a time to event endpoint, a reference time – either historical or comparator – is needed. The FDA appears to prefer a comparator to a historical reference time. The FDA official said "Every measurement you take has a certain fuzziness about it. That means the measurements aren't exactly what they say. What you are looking at is likely but could be within a range (a confidence level), which is determined by how the numbers are calculated, and that is dependent on how many people are in the trial and how big the effect is. The bigger the effect, the more accurate the results tend to be. The more people, the more accurate the trial is. But 90% of what we see is a modest number of people in the trial and a modest effect. Almost every study done has a confidence interval around the results that sometimes can be pretty big. To have the most confidence in the results, you want the comparator to be done at exactly the same time as what you are testing – that is, another arm to the trial, not a historical reference. There are two reasons for this:

(1) the inherent in measuring everything, and

(2) what's called the secular effect in medicine, meaning care changes over time. You can't compare a 10-year-old car to a modern car. All kinds of little things change that add up in the health care picture, such as better scanning, better detection, better nutrition, etc."

A prominent oncologist warned that using TTP can raise the question of whether patients enrolled in the trial were cherry-picked. He said, "As we get better at early diagnoses, take more account of quality of life and subdivide tumors, regulators have issues with all of that. You could argue that we are cherry-picking. If a drug is active in 10% of breast cancer patients with a 50% (improvement) rate, but in overall breast cancer is not a good drug, regulators will say, 'Show me that the group of patients in which it worked is not cherry-picking.' I don't think you will be allowed to cherry-pick – unless there is a very, very clear relationship to sensitivity and your drug. The specificity of the molecular mechanism is one way you might work your way thru that debate."

Three factors make TTP a difficult endpoint:

1. **Bias.** This generally can be avoided with a blinded trial, but blinding a trial can be difficult if the study drug causes obvious side effects or has to be given in a clearly different manner. For example, patients with a study drug might all lose their hair, or the study drug might cause injection site changes or be a different color from the comparator. The FDA official said, "If you are doing a study where you are trying to measure TTP, you should be comparing simultaneously in a randomized population to a defined comparator or you will get bias and uncertainty. You can try to minimize uncertainty with large numbers of patients, but you also have to minimize bias…Blinding is really hard in oncology, but even if you could perfectly blind a trial, it doesn't necessarily make it easier to use tumor TTP.

2. **Patient selection.** The FDA official said, "There is a well-known cancer researcher who gets results no one can duplicate because the only patients he treats are Olympic athletes with a little bit of cancer. So his results don't necessarily apply to the general public."

3. **Measurement interpretation**. Measuring tumor progression can vary depending on who the reviewer is and how the measurement is made. The FDA official said, "How you define progression depends on:

   (a) the size of the effect

   (b) the precision and timing of the measurements

   (c) the technique of how you are measuring tumors – x-ray, MRI, CT, questionnaires, symptom scales, etc. Measuring a three dimensional, irregularly shaped tumor on an image is not necessarily that precise. And, if something is only measured at the time of the x-ray or image, that affects the interpretation of the measurement. Even if you take a measurement every six weeks (which is frequent) or once every three months (which is more normal), how do you know the tumor didn't appear four days before the most recent scan? You don't know when the tumor appears, so assignment of time is very imprecise.

TTP is a less precise endpoint than survival. The FDA official explained, "One of the reasons survival gives you a lot of confidence is that it's hard to fudge…Say Drug A has six month TTP based on tumor size, and the comparator has three months, but your uncertainty is ±three months. Then Drug A and the comparator could essentially overlap, and you are not really sure what you are seeing. I've seen and reviewed applications with a clear statistically significant difference in

TTP where progression was measured as tumor size, and there was no difference in survival. In fact, survival might have been worse, not statistically worse, but numerically worse in the arm with the better TTP. The explanation for this is that the way progression is defined gives you a degree of uncertainty that can lead to a potentially false conclusion about the net effect of the drug. So, if progression is solving a problem for the patient, that problem-solving can be meaningful for the patient, but if it is a lab finding or an image finding, then what are you really doing? It is a nice experiment, but are you really showing any clinically significant effect?"

Another issue in analyzing TTP trials is that patients in Phase III trials tend to be less sick than the patients used in Phase I and sometimes Phase II trials. A researcher said, "Often, the worst, the sickest, patients are used in Phase I trials. Then, patients who are a little less sick are used in Phase II, and in Phase III they take patients even earlier, so median TTP can't be compared from Phase I to II to III trials. It is very common for TTP to improve from Phase I to Phase III." The FDA official said, "You probably get better results in a Phase II trial than a Phase I because patients tend to the be most heavily treated and have the worst prognosis in a Phase I. But you may or may not see a difference between Phase II and Phase III patients, though you can do a 20-patient Phase II study and have some spectacular results that get everyone excited, and then the next group of people who get the same treatment at a different site do not have the same results."

Thus, TTP can be a very fuzzy endpoint, but sometimes it is chosen when using a placebo arm is not an option. A researcher said, "Survival is a black and white endpoint. Median time to death drifts dramatically in trials. TTP is a really fuzzy endpoint. A randomized, double-blind, placebo-controlled trial is the most precise and definitive way to know if a drug is working, but the problem is whether you can ethically do that in a metastatic cancer trial. The compromise is a TTP trial. But the FDA will say that TTP is murky and that it does not make clear the drug's benefit. There also has been a lot of bad data in TTP trials. If you use TTP improperly – if it is non-randomized, not blinded in assessment or has bad criteria for progression – then it is a very hazardous area. I understand the mistrust the FDA has, but it hasn't defined what will instill trust. People do TTP trials because patients will participate, and they get a biological answer but not necessary FDA approvability."

Two sources believe FDA should make its position clearer on the use of TTP endpoints: "The FDA needs to decide what it will permit in terms of size and methodology in order to believe in TTP trials. There is a debate whether the Avastin design will convince the FDA. My own Avastin trial in kidney cancer did not impress the FDA at all, even though the p value was <.001. The agency was completely unimpressed because the degree of difference in the two arms was small. The reason the difference was small was that we set precise and small criteria for progression. We allowed a 25% increase in any tumor. That is a very small difference, but it was enough for us to call it progression."

However, the FDA has no immediate plans to issue comprehensive guidelines on the use of TTP endpoints. The FDA official said, "The TTP endpoint is always being debated, but progression can be hard to define in a consistent and general way. In oncology, the answer is: Don't anticipate a document that would be definitive in the near future." Asked for guidance on appropriate trial size and methodology for using TTP in a way that would enable the FDA to have some confidence in the results, the FDA official said, "It depends on how progression is defined. If progression is highly reliable, and you see a direct patient benefit or a reliable surrogate for patient benefit, then it makes sense. The trial size depends on how big an effect you see or anticipate seeing and how you measure it. The reality is most anti-cancer therapies have only a modest effect, and so the regulations say you need to replicate your findings, which is why it says trials, plural – because any one trial could be a fluke. We interpret that many ways. If it is a very large trial with many sites, particularly a world-wide trial, and all the results are consistent, then maybe one trial could suffice."

Could a TTP trial be used as a confirmatory trial? One researcher didn't think so, but the FDA official insisted that it could. A researcher said, "From the impression I got from the FDA, it would not consider my trial at all in the approval process, not just that they would weight it less – they wouldn't weight it at all."

TTP may have the most usefulness in neurologic symptom trials, though it could be used in any area. The senior FDA official said, "TTP makes a lot of sense when you are talking of a treatment that is supposed to alleviate symptoms. There are different ways of looking at a clinical trial. The therapy is trying to solve a problem for a patient – the patient has a problem, seeks advice, and someone makes a recommendation on how to solve that problem. There is a goal with a lot of life-threatening diseases, which is to get rid of the disease, but that is not always possible. So, there are two choices, which are not mutually exclusive. If the patient can't be cured, will the treatment prolong life and/or will it make whatever problem the patient is having less or even go away? You could call it quality of life, but that is a harder concept. The more contemporary approach is patient-reported outcomes. No one quite knows what quality of life is. There actually is huge interest in quality of medical care, looking at it like a process, a manufacturing process. Quality manufacturing is pretty well understood. Quality service is a little harder to understand but can be measured – you can decide what annoys people and what impacts negatively on people and try to minimize those. But when you say, 'What is the global

quality of life?' Who knows how to answer that? There is a lot of interest in the topic, but to date, I don't think anyone has gotten a product approved for marketing on the basis of quality of life per se and certainly not in cancer."

Endpoints in cancer trials are obviously a little different than for some other conditions. The FDA official said, "We have approved cancer therapies because of a decrease in pain, alleviation of symptoms such as if you have a tumor and can't swallow or breathe, or a tumor that is interfering with something important, and then you get a therapy and are able to do breathe, swallow or do that thing. That is solving a problem, and my personal – not official – opinion is that is problem-solving, not life-solving. Using TTP is similar. It means trying to avoid having a problem, trying to delay something getting worse. In that framework, it is relatively easy to understand the neurologic use of TTP: You either have a problem or you are likely to get a problem, and you want to delay getting it."

Researchers said several breast cancer drugs have been approved in the past based on a primary endpoint of TTP, including Genentech's Herceptin (trastuzumab), Novartis' Femara (letrozole), AstraZeneca's Nolvadex (tamoxifen) and Arimidex (anastrozole). However, an FDA official said, "Something can be approved based on TTP, but it must clearly demonstrate patient benefit. A single-arm study with no comparator could be open to bias and uncertainty, but a well-designed, randomized, controlled trial could support approval."

What is the FDA's current position on the use of TTP in oncology trials? The FDA official said, "Any time you talk about TTP, unless it is a slam dunk with miraculous findings where you give everyone the drug and they have dramatic results that last the rest of their lives, you would be discussing randomized controlled studies. For example, in Lou Gehrig's disease (ALS, amyotrophic lateral sclerosis), if you had a TTP endpoint that measured the time before a patient lost the ability to speak, how would you know when the patient would have lost the ability to speak without the drug? You have to do a comparison."

Could a TTP trial be stopped because of "outstanding" efficacy? The FDA official said, "Why not? But you have to have confidence that it is meaningful. I can imagine a scenario where it would make lot of sense. Say you are comparing Drug A to a standard comparator, and measuring a symptom, to pick something clinically obvious, such as hair loss, with a goal of delaying hair loss. If, with Drug A half the patients lose their hair at one year, and half the patients on the standard comparator lose their hair at six months, and you've done the analysis and said, 'If the median TTP for the comparator was six months, and on Drug A it was a year, that is your clinical benefit.' We also have a program called accelerated approval – the approval process is not necessarily accelerated, but the drug cycle is accelerated – so we will in some instances, review an application before clinical benefit has been shown, and, depending on how you define TTP, that may or may not lead to marketing approval. If you are measuring a tumor biomarker, then you don't have the clinical benefit; you are measuring something else. You don't know if that is a good surrogate for the clinical benefit or not. It might correlate, but under Sub Part H for accelerated approval, you may be able to use that biomarker to file an application if it is considered reasonably likely to predict clinical benefit. Most of these applications have been taken to advisory committees to provide a public discussion and to solicit input on the use of the surrogate."

The FDA's message to researchers: "If you are doing a well-designed trial and your definition of progression is something clinically meaningful and can be measured with some degree of precision, that is the kind of study we are inviting people to perform." ¨