TRENDS-in-MEDICINE

April 2013

by D. Woods

SUMMARY

The FDA's Dr. Robert Temple offered advice on strategies that companies can use to "enrich" the patient population in their trials so that the trials don't have to be huge. Basically, he said that enrichment – noise reduction, prognostic enrichment, or predictive enrichment – has become a critical component of most clinical trials, but it needs to be done carefully.

Trends-in-Medicine
Stephen Snyder, *Publisher*
2731 N.E. Pinecrest Lakes Blvd.
Jensen Beach, FL 34957
772-334-7409
Fax 772-334-0856
www.trends-in-medicine.com
TrendsInMedicine@aol.com

FDA VIEW OF ENRICHMENT STRATEGIES FOR CLINICAL TRIALS

White Oak, MD
March 25, 2013

Sometimes the only way to determine the safety or efficacy of a drug is to do a very, very large clinical trial, but there are some strategies that companies can use to "enrich" the patient population in their trials so that they don't have to be as big. In an effort to help companies know how and when to use these approaches, the FDA issued draft guidance in December 2012 on enrichment strategies for clinical trials. Robert Temple, MD, deputy center director for clinical science in the FDA's Center for Drug Evaluation and Research (CDER), discussed these strategies in a webinar recently with industry (and a few reporters). Basically, he said that enrichment has become a critical component of most clinical trials, but it needs to be done carefully.

Dr. Temple outlined three kinds of enrichment that can be used in clinical trials to find a study population in which the effectiveness of a drug can be shown – noise reduction, prognostic enrichment, and predictive enrichment. He said, "A major contributor to efficacy includes the likelihood of showing a drug effect, if there is one, by choosing the right patients for the trials. Enrichment is the effort to make sure that people in the trial have the disease that we are studying, don't respond too well to placebo, and do not have conditions which would obscure benefit."

He added, "Enrichment is the prospective use of any patient characteristic – demographic, pathophysiologic, historical, genetic, etc., to select a study population in which detection of a drug effect is more likely than it would be in an unselected population. Doing this increases study power, facilitating proof-of-principle. However, it can leave concern about the generalizability of the result. This raises the additional question: How much data do we need before and after approval in the non-selected group?"

Dr. Temple said the issue of generalizability is unavoidable, "Sometimes you can easily define the selection criterion – i.e., a genetic defect. But with empiric designs – for example, doing studies in people who respond to an open screen – there really is no way to identify the responder population; you just know that there is one. In some cases, the remedy is to use the designs early. To show unequivocal drug effect, don't make the enrichment study the only study, at least not usually, and be aware of what you've done and don't hide it or overstate results. Searching for who the responders are…is highly worthwhile and improves the likelihood that you will show effect."

Dr. Temple said that the guidance is primarily focused on studies intended to demonstrate effectiveness, but it also is pertinent to safety studies. For example, he said that with oral

anti-diabetic drugs trying to reduce cardiovascular (CV) risk, "We recognized that to have any chance of succeeding…you need to include high-risk patients…So, more and more studies are including people who are older, who have had diabetes for a longer time, and who have a history of CV risk. To succeed in a safety study, you have to do that. If a drug class has an effect that bothers a lot of people in another way, a way to find out if a new drug lacks the effect is to do a study in people who have had the effect with a previous drug. This was used with losartan [Merck's Cozaar], and it worked."

The three kinds of enrichment:

■ **Noise reduction** is a way to include people whose disease is stable and who can be measured precisely. If you can remove people with a significant placebo response, then the difference between active treatment and placebo will be obvious. Dr. Temple listed some elements of noise reduction – for decreasing heterogeneity – that, in general, do not raise questions of generalizability:

- Define entry criteria carefully to make sure patients have the disease being studied.
- Find prospectively likely compliers (VA hypertension studies, Physicians Health Study, for example).
- Choose people who will not drop out.
- Eliminate placebo responders in a lead in period.
- Eliminate people who give inconsistent treadmill results in heart failure or angina trials, or whose blood pressure is unstable.
- Eliminate people with diseases likely to lead to early death.
- Eliminate people on drugs with the same effect as test drug.

■ **Prognostic enrichment.** This mostly applies to studies where the goal is to show that a drug reduces a bad outcome, such as heart attack or death. In order to succeed, the trial population must have a reasonable number of those types of events. If the group is too healthy, it won't have enough events, and the drug will look ineffective. In studies of drugs to treat symptoms, people often try to do the early studies of a drug in a population that is reasonably sick, so that there is something to improve. Dr. Temple said that choosing high-risk patients "has therapeutic implications, and you may be able to do a smaller study."

Choices could be based on patient characteristics (patho-physiologic, proteomic/genomic), or be empiric, based on patient history of response (i.e., tumor response on some radiographic measure), or response to a test drug.

Dr. Temple described his favorite example of prognostic enrichment: the CONSENSUS trial of the ACE inhibitor enalapril in NYHA Class IV patients. Using NYHA class is a simple way of ranking the severity of heart failure. He said, "Later studies of less sick people needed thousands of patients. The fact that these people were very high risk made a big difference in the trial…Similarly, the simvastatin trial – the first lipid outcome trial – was done in a post-MI, very high cholesterol population, so having a lot of events really helps you – if the drug works, of course."

Dr. Temple said that there is always some concern about people who aren't in the selected population, "If you take steps to reduce heterogeneity, no one is concerned. If you've gone through prognostic enrichment, you'd have to look at a lot of labels. We tend to put the exact nature of the study and how the drug was studied in the clinical trial section [of labels]. In a predictive enrichment, you generally get a claim for the population you have studied, but there is always some uncertainty about the perfection of the marker – how good is it."

■ **Predictive enrichment.** This means trying to find people at high risk of having the disease. Epidemiologic risk factors for CV outcomes include: recent events (e.g., acute MI or stroke), history of angina, cholesterol, blood pressure, diabetes, elevated CRP, family history, gender, race, age, etc. Individual measurements/history also can be used, such as tumor histology, genetic, echo findings, etc.

Dr. Temple said, "In one way or another, it is routine to try to find people at high risk. This is common in oncology and cardiovascular medicine," such as:

- Prevention of breast or ovarian cancer in people at high risk. Tamoxifen prevented contralateral breast tumors in an adjuvant setting (very high risk). It was then studied in people with a more general high risk. This was needed (a) to have enough endpoints to detect possible effect and (b) because of concerns about toxicity. It was labeled for the group studied. There was no reason to expect a larger percentage effect because you wouldn't be exposing lower-risk people to toxicity. Dr. Temple talked about tumor genomics and different gene expression profiling approaches intended to predict breast cancer recurrence rates.
- Cardiovascular outcome studies of lipid-lowering agents – patients with a history of acute myocardial infarction, very high LDL, low HDL, or elevated CRP. He said, "It

has long been routine to choose patients at risk for outcome studies. These are often called secondary prevention or post-AMI or post-stroke, and it makes a huge difference."

- Studies of antiplatelet therapies in angioplasty patients.

He said, "People have been looking at these for a long time. Getting in early in Alzheimer's disease is something on many people's mind – how to identify people at high risk…There is always a benefit:risk question in lower-risk patients, but you have already established efficacy in high-risk patients, which in my view puts everybody ahead."

Dr. Temple also described the recent JUPITER study of AstraZeneca's Crestor (rosuvastatin), saying, "These people had to have elevated CRP, which is at least a predictor of having a heart attack. It was a large trial – almost 18,000 people – and these were basically not very sick people. The endpoint was the first major CV event, defined fairly broadly…The rate of primary endpoints was fairly low, 1.36 per 100 patient-years in the untreated population, but I'm sure that was higher than it would have been if they had taken an unselected population."

Dr. Temple said that there is tremendous interest in high-risk patients in trials like Alzheimer's disease or particular cancers.

He said predictive enrichment "is lighting everybody's fire. It is identifying people who will respond to the treatment and then studying them. This enormously enhances the power of the study. It is especially critical if responders are a small percentage of people who have the condition. Selection can be based on tumor receptors, pathophysiology, or it can be empiric. Pathophysiology can be something like hypertension…We study antibiotics in bacterial infections. A well-established genetically determined difference could be the basis for a pathophysiologically-selected population. Many tumor genetic or surface markers can be used."

In the selection of likely responders, even if the pathophysiology is unclear, likely responders could be identified by an initial short-term response, which is an empiric approach. For example:

- The CAST trial was done in patients who had a 70% reduction in ventricular premature beats (VPBs), and only "responders" were randomized.
- Trials of topical nitrates were done only in people with blood pressure or angina response to sublingual nitroglycerin.
- Anti-arrhythmics were developed by open screening for response, then randomizing the responders.
- Every randomized withdrawal study has this characteristic.
- History of response to a class.

Dr. Temple said, "I've long suggested that people do dose response studies, and, to my knowledge, no one has done that…You can also select likely responders based on some sort of biomarker, like a tumor that shows early metabolic effect on pet scan, one that shows early response on blood measure (like PSA), a tumor that doesn't grow over an n-week period, only patients with LDL effect >n, or some other less studied lipid, only patients with CRP response >x. There are lots of possibilities, and you gain a lot."

If a trial is done entirely in a marker group, Dr. Temple said that one question that always arises is whether it is efficient but gives no information on the omitted patients, "The guidance urges repeatedly that unless there is no real chance of an effect in marker-negative patients, some negative patients should be included because they may have some response, and their data can be used to refine the marker cutoff. It would still be possible to make the primary endpoint the effect in the enriched stratum."

He added, "The guidance spends a fair amount of time describing empiric approaches. You don't necessarily understand why it works in the group, but it appears to in an early look. Beta blocker heart failure studies were also done sort of in this way, but what was screened was tolerability, so the trials – most of them – were only done in people who could tolerate it. It seems that the effect kind of overstates the effectiveness in the unselected population, but it helps you do the study."

Dr. Temple mentioned the interesting wide variety of pathophysiologic or genetic characteristics, "You could do a study only in people who make the active metabolite for clopidogrel. You could study only people whose tumor takes up a drug. You can look at effects on tumor metabolism, and there is tremendous interest these days in looking for proteomic or genetic markers that predict response. Hepatitis drugs have used it, and that clearly is the wave of the future in many ways."

Predictive enrichment using genomic proteomic selection has mainly been in oncology, but Dr. Temple also pointed to Vertex's Kalydeco (ivacaftor), a cystic fibrosis (CF) drug that was found to be useful in a small fraction (4%) of all CF patients, adding, "The use of [predictive enrichment] is clearly spreading."

Randomized withdrawal is a way to find out who responders are. The method is efficient and ethical, which is an attractive point in pediatrics. Dr. Temple said, "[In angina trials] people would get open treatment with the test drug, and people who responded could be randomized to drug or placebo. The endpoint could be time-to-failure, in which case people don't have to feel miserable for very long, or it could be a conventional measure…You are finding out who the responders are. Also, they often need no recruitment and this is important in rare diseases where finding patients is difficult. This can be the way to get all the people you need. All antidepressants are tested for maintenance effects using this kind of design. People who respond well for ≥3 months are randomized to therapy or ending of therapy, and depression is the endpoint…These randomized withdrawal maintenance studies never fail."

Dr. Temple said that one question about predictive enrichment is whether a treatment might work for people for whom the previous therapy did not work, "Studies in non-responders, randomized to a new drug and the failed drug – I am only aware of four studies which did this, including clozapine [an antipsychotic used to treat schizophrenia] and a beta blocker… You can also do it in people who do not tolerate a drug… There is a long section in the guidance about what to watch out for in considering predictive enrichment designs and the properties – advantageous or not – of specific designs."

Some things he said to keep in mind:

- ***Performance characteristics of the selection criteria.*** When a test (genomic, proteomic) is used to choose patients, you need to know test precision.
- ***When to develop the classifier.*** Ideally, early studies would enter a broad range, and evolving data would help choose a cutoff.
- ***Who to include.*** It could be only enrichment population patients or all patients, but analyze only those with the marker as the primary endpoint.
- If there is ***no*** effect in the marker-negative group, you can screen all patients and randomize only marker-positive patients.
- It supports the effect for an enriched population, but it overstates the effect for the non-responder population.
- It is suitable when there is little or no chance that the marker-negatives would respond, and the labeling would have to say who was studied.
- If you show the disease exists where there is no drug, the FDA "will likely approve that."
- If you do a stratified study – in which both groups, positive and negative, get randomized to drug and placebo – it could be all marker-positive or a third or a quarter marker-negative. Dr. Temple said, "We don't have a position on that. Sometimes you can't tell at randomization whether someone is positive or negative, and in that case you have to randomize everybody."

Questions and answers

Asked if the FDA would consider a drug for approval if two prior trials suggested predictive markers in a subset of the population, Dr. Temple said, "It is hard if you have one definitive study in an enriched population and two prior ones showing good predictive markers, might we approve it? You could describe the IDEAL trial as that kind of case – definitive in a black population – and led to approval, but that was because we thought we were halfway there with the veteran studies in a mixed population."

Asked about oncology and predicting response in one type of cancer where a biomarker has not been validated, Dr. Temple said, "We have seen enough examples when a particular marker was predictive in one tumor and not another, so you have to provide data that shows the marker is predictive in the tumor of interest…Usually the success of the predictive marker is specific to a particular tumor, sometimes even the stage of the tumor."

Asked whether a smaller size would be sufficient for a safety assessment, Dr. Temple said, "We have written what we expect for safety in a broad and general way…The magic number is something like 1,500 patients. If you treat a disease that has no treatment, and an enrichment design allows you to do that, you see approvals in oncology with, at most, a couple hundred patients, and we are willing to do that in something very important. If it were a widespread disease, we would expect additional data, but it's got to be case-by-case. It depends on the magnitude of improvement."

As for devices, he said, "There is no doubt that some devices are studied in people with relatively advanced disease where you think that getting the device makes a very big difference in likelihood of outcome. But I don't know enough to answer that question."

Asked if the FDA had sponsors in recent years submit Phase III trials in pediatrics, Dr. Temple said, "In fact, they are regularly used in pediatric hypertension studies that include a placebo group. So, they randomize to several doses of the drug. If the high dose is better than the low dose, you know it works. Sometimes all the doses have the same effect, and so you know you didn't

pick the right [doses]. It has been used in that very setting, and that focuses on the fact that, depending on the endpoint, sometimes with a randomized withdrawal study, you can find out when someone isn't being treated well, very, very quickly. And you can have a withdrawal criteria."

Asked about randomized withdrawal studies, Dr. Temple said, "Whether a randomized withdrawal study works depends on the mechanism of the drug. If it has a prompt effect, a withdrawal study works fine. If it has prolonged effect, you may not know how to do a randomized withdrawal study. You may have to wait six months. Not that you couldn't, but I'm not sure how attractive that would be for people…I don't think that these would be Phase I studies. These are controlled trials we're talking about. It may be very sensible to do an early study in a highly enriched population to get some evidence that the drug really works. That's what the anti-arrhythmic studies did. They screened a population for the ability to respond…and randomized them into these trials which were very efficient, very successful, and not very large, so it made a lot of sense. Once you know the drug works, you might want to choose a broader population to see how it works. I'd say Phase II, but they are also very useful in Phase III for the definitive trial."

Asked about situations where two studies are strongly suggestive of a phenotypic marker, Dr. Temple said, "I don't think I have an answer. Ordinarily, when you go scratching through the data already collected, we are a little worried that multiplicity will allow you to reach a conclusion. The suggestion here is that you've got studies showing that they really seem to show a marker predicts effectiveness. You want to know how sensible/plausible the mechanism is…I wouldn't rule it out. A very important factor is: Did the study win overall? So, I'm assuming that the two well-controlled studies actually showed an effect of the drug, and it seemed to be particularly present in the subset…We sometimes do pay attention to that, especially if the study wins. It is very valuable information regarding who gets the drug."

Asked if the FDA would consider multiple cancer confirmatory trials based on a common genetic aberration, Dr. Temple said, "I'm not prepared to answer that. I have to talk to the oncologists. As a general rule, more and more one is thinking of cancers as different and not so much all the same…And we have seen phenotypic and even genotypic markers be predictive in one cancer and non-predictive in another cancer."

Asked for an example of a population in which treatment was initiated after completing doses, Dr. Temple said, "I don't think that there is necessarily a problem if you start people on standard of care and give that for a month and then randomized to addition of treatment plus another treatment. I'm not sure I know of any cases where that's what you do. But I don't see an inherent problem in that…Certainly, there are trials where people failing on standard of care for some period are the people you put in the trial. That is an acceptable study design. But it means the overall rate of response in control is relatively low…You would continue and then add the new drug. That could be acceptable, but I'd need the details."

Asked about a situation where an imaging agent a sponsor wanted to use in a trial was not FDA approved, Dr. Temple said, "We have a basic conception that if a diagnostic test is a necessary part of the treatment approach with a drug that it needs to get through CDRH [the FDA's Center for Devices and Radiological Health) and be approved for that use. Our guidance says almost always. But if you have done something spectacular and have saved people's lives or it looks like you could and the imaging device isn't approved, we would probably approve it while they work on the imaging. Some imaging modalities are not CDRH approved, but we can imagine if something were important enough it could work on that basis."