

NCI 2007 PRE-OP THERAPY IN BREAST CANCER
38_SESSION 6_7 - BERRY

DR. CLIFFORD HUDIS: Our very next speaker is Don Berry. He's the Chair of the Department of Biostats and Applied Math at MD Anderson. He is, as well, a statistician on the breast committee of the CALGB. Don is going to talk about statistical considerations in preop trials.

DR. DON BERRY: Thank you. There's some people who think statistics is the most exciting part of a conference such as this. (Laughter) Unfortunately, none of them are here. (Laughter)

My charge was to talk about statistical considerations, and I wanted to dispense with my charge in one slide, and then talk about something I hope will be mildly exciting, at least.

The sample size considerations for a dichotomous endpoint are standard, whatever the endpoint. So, if you want to improve a response rate from 20 percent to 40 percent, it doesn't matter whether you're talking about metastatic disease or pCR -- the statistical issues are the same.

In the metastatic setting, the response is, of course, somewhat different -- much more problematic, because we don't measure at a particular time. We look at, you know, after six weeks, after twelve weeks, etc. And at some point, we register that there's a response, and time to response. Or the fact that you haven't responded, but haven't progressed, by six months is important. We never consider such things. We probably should. Indeed, in the metastatic setting, response is just one part of the continuum of progression and stable disease. Some of the same issues arise in the neoadjuvant setting.

So, here's my outline. I want to tell you about adjuvant trials and ask the question, are they still viable? They're getting huge, for good reason. It raises the issue of, isn't there something better? Like, for example, neoadjuvant trials.

PRE-OP THERAPY IN BREAST CANCER
38_SESSION 6_7 - BERRY

I want to discuss, with deference to Antonio [Wolff] and others, the issue of surrogacy. I will define “surrogacy” from a regulatory setting. And, after I do that, I hope that we never use the word again. (Laughter)

I'll talk about modeling. You know, something that we don't do very well in cancer, and indeed in medical research generally, is model the patient. You know, what happens to the patient over time?

We ought to be looking at pCR. We ought to be looking at imaging, doing biopsies when and if we can. We ought to be looking at the relationship between the early endpoints and the later endpoints. And I'll give you an example of that. And, finally, a little on the fine-tuning of pCR.

So, here are... CALGB node-positive studies in [since] 1975.

Comparison -- among other things -- a comparison of CMF with CMFVP [CALGB-7581]. 800 some patients.

Nine hundred some patients in a study designed five years later [CALGB-8082]. A very complicated design comparing, at some course in the treatment, CMFVP with VATH. You know, five drugs versus four other drugs, in retrospect, we didn't learn much about individual drugs. But that's another story.

CALGB-8541, designed five years later. Sample size getting bigger.

PRE-OP THERAPY IN BREAST CANCER
38_SESSION 6_7 - BERRY

Eight years later, [CALGB-]9344. This is the Taxol / three doses of Adriamycin study. Sample size even bigger. Now, these are node-positive patients, where you expect a reasonable number of events.

And, in fact, we've come to learn that whatever number of events we expect, we don't get. And for good reason: the disease is changing, the treatment is changing. The prognosis is getting better and better and better, for those who are detected.

We targeted 1,800 events in this study. There were interim analyses after 450; 900; and 1,350. That study was supposed to close and be announced three years after the last patient accrued, in 1997. It is now 10 years after that, and we still are not at the third interim analysis point. Luckily for the profession and patients, we announced the data a long time ago.

That's disease-free survival. This is survival. The same sort of thing. Not quite as extreme, but extreme nonetheless, showing an increasing benefit over time, because of screening mammography, stage shift associated with it -- even within Stage II disease -- the introduction of tamoxifen, and, of course, the improvements in chemotherapy. So, things are getting better and better.

In the ATAC trial -- a 9,000-patient trial that we've heard about here. And... something bad happened on the way to the thing... [Much of the slide turns out to be blank]. So, I guess you're not going to get that.

So, let me tell you what that is. If you look at the picture from the original publication, it shows three curves that are that far apart, with a highly statistically significant difference. And, of course, the disease-free survival curves are very high -- above 90 percent for most of the duration. They then show... they magnified it so that you, you know, could

PRE-OP THERAPY IN BREAST CANCER
38_SESSION 6_7 - BERRY

see the difference between the curves. And what this thing up here is -- the 10 days -- was the difference over the 3.5-year period between anastrozole and tamoxifen. Several points about that -- the main one is that the sample size is huge and that we're looking for smaller and smaller benefits.

So, is there potential for more sensitive and earlier comparisons? And the neoadjuvant setting provides a very neat example of that. So, I want to revisit Aman Buzdar's trial that he told you about, just to make one point. This was... well, he already explained it. The design was 164 patients, randomized equally. No interim analyses. But the data monitoring committee analyzed the data after 34 patients. I think Aman showed you something like 39 patients.

But after 34 patients, these were the response rates -- pCR rates. We said, especially in view of the slow accrual -- it was accruing at about two patients per month -- "We know the answer." We did a calculation saying, given these data, what is the probability that, after 164 patients, we'll have statistical significance? And the answer was 95 percent. We said, this is enough -- we know the answer. Of course, the precision associated with the estimates is nowhere near what it would be eventually. But we don't want to wait another eight or nine years to publish these results. And so, the data monitoring committee opened it up. And it was presented at ASCO and [in] *JCO*.

So, here's the chronology -- thousands of patients on trastuzumab treated in the metastatic setting. And 12,000 patients treated in the adjuvant setting. And the teeny Buzdar trial in between, with 34 patients -- admittedly, a different endpoint; but kind of a link between the metastatic endpoints and the adjuvant endpoints and, of course, small.

What about pCR? It's wonderful statistically. It's wonderful because -- for several reasons. One is that it's at a fixed time. So it's not like in metastatic disease where we

PRE-OP THERAPY IN BREAST CANCER
38_SESSION 6_7 - BERRY

say a patient responded at some time during the course of therapy. It happens at a fixed time, at least within treatment group. It's early. And the early aspect means that we can take advantage of what we learned. So, it's wonderful for being adaptive. It's wonderful for building a sleek design that addresses questions that arise, that validates them and gets on with looking at other things.

pCR should be fine-tuned along the lines of Fraser Symmans -- reducing the dichotomy into, you know, how much of a CR is it? Or how much of a response is it?

But is it [pCR] a surrogate for anything that's of... clinically relevant? So, the regulatory attitude towards surrogacy is the Prentice criterion. The Prentice criterion is, "a response variable for which a test of the null hypothesis of no relation to THE TREATMENT GROUPS" - and stress that - "under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint."

Now, only a statistician can write that. (Laughter) And if I were to ask you, what the hell does it mean? (Laughter) Roughly, it means that if you tell me the pCR difference between the two, you don't have to tell me anything else. You don't have to tell me what the disease-free survival was -- that I know the answers. I know whether the treatment is effective. Roughly speaking, it says that they're -- the clinical endpoint and pCR -- is perfectly correlated. That's not exactly true, of course. The requirement is not quite that strong; but it's pretty strong.

It's a very high hurdle. pCR clearly does not qualify. And we've seen that in the conference so far. But it's useful nonetheless. And I want to give you an example of its usefulness: this is using neoadjuvant therapy in drug development. And it's being adaptive -- so, there are two things that are going on. One is, we're doing something revolutionary in medicine -- medical research. We look at the data. And we adapt to

PRE-OP THERAPY IN BREAST CANCER
38_SESSION 6_7 - BERRY

what we're seeing. The other is, in the neoadjuvant setting, as opposed to the adjuvant setting, we've got something to look at. Namely, whether the patient experience -- whether the tumor was a pathologic complete response.

So, this is, in a way, a rather simple adaptive design. I didn't want to burden you with the more complicated kinds of things that we do. This is a seamless phase II [II/III?] design, where the phase II aspect is trying to see if there's a signal -- pCR mostly; but the primary endpoint is disease-free survival, and we're modeling the relationship between the two.

So, the phase 2 aspect is going to be a small number of centers accruing. And the only difference between phase II and phase III is that, in phase III, we expand to a larger number of centers. All the while, we're observing the relationship between pCR and disease-free survival BY TREATMENT GROUP, and we're validating the relationship that we see. One of the problems about surrogacy is it has to be true for all treatments. What we're doing is we're focusing on a particular -- and we know that that's certainly not even close to being true with pCR.

So, I say standard approach. Of course, it's not the standard approach. The standard approach is, we do phase II trials in metastatic disease and then take it into other metastatic disease populations and primary cancers. But we're moving toward -- and we certainly could do what most of drug development does -- namely, have an early endpoint in the phase II part, and then a later, more clinically relevant endpoint in the phase III part.

It's dumb. It's incredibly dumb. We find that we have an effect on the phase II endpoint, and no effect on the phase III end point. Surprise! What we ought to be doing is looking

PRE-OP THERAPY IN BREAST CANCER
38_SESSION 6_7 - BERRY

at everything at the same time and trying to understand the more important question, is there an effect on the clinically important endpoint? In this case, disease-free survival.

So, in the standard approach, if you observe a sufficient pCR rate, you may say “Yahoo! Let’s go to phase III.” And if you observe a disease-free survival advantage, you market your drug and everybody’s happy. If you don’t, then you, you know, find another job. If there’s a low pCR rate, you stop accrual.

And in between, there’s this white space for -- I don’t know -- analyzing the data, deciding what you’re going to do. The total development time from phase II to the end is seven years -- maybe more than that.

In a seamless phase II/III -- that I’m going to describe and there are particular assumptions that I’ve made that I won’t tell you -- but, relatively, the same sort of thing will apply. That is, if you’re looking at a smaller advantage, the same comparisons will apply. So, in the seamless phases, you start out at a few centers -- like 15 patients per month -- comparing equally randomized experimental and control. If the predictive probabilities look good -- so what you do is you say, “Here I am today. What is the probability that experimental therapy will be shown to be better than control therapy?”

And if that gets to be, you know, moderate -- not huge, but moderate -- so that there’s a reason to continue, then you expand into phase III, and let’s say, start accruing, you open up other centers -- presumably, the paths have been greased, the IRB paths have been greased in advance -- and you accrue, in this example, 60 patients per month. The maximum sample size is 180. This is a single trial. All patients in the trial count for the final analysis, which is disease-free survival.

PRE-OP THERAPY IN BREAST CANCER
38_SESSION 6_7 - BERRY

You can stop early. If the predictive probabilities and statistical significance are low, you stop for futility. Presumably, that would mean that you're not seeing much, if anything, in pCR rate.

And you're getting some information, not much, but some information, about how a pCR rate relates to disease-free survival. That information will be treatment-specific. And the most important patients for that are the phase II patients -- the ones that came in early -- because there's longer follow-up.

Kind of an irony is that we run phase II on patients who have less follow-up than would be possible for phase... [corrects himself] ...we run phase III on patients that have less follow-up than the better, longer follow-up in phase II.

We use predictive probabilities to switch to phase III. If, as I said, they look good, we can stop accrual for superiority and continue follow-up for some period of time. Or for futility.

So, in comparisons -- I want to compare this design with the standard design or conventional design. And I use two conventional designs. One has four analyses total -- three interim analyses and a final analysis. The other has 18. And the reason I chose 18 is because it's the maximum number in the Bayesian analysis. We look every two months at the data and do predictions and decide whether we're done yet or we want to expand accrual. And, of course, we can be done by being a positive or a negative.

And all three of these designs have the same significance level and the same statistical power. This is the expected number of patients in the Bayesian adaptive design that's using pCR rate, modeling with disease-free survival.

PRE-OP THERAPY IN BREAST CANCER
38_SESSION 6_7 - BERRY

The two conventional treatments UNDER THE NULL HYPOTHESIS.

Now, I say, THE null hypothesis. There are lots of null hypotheses. They all look like this. One null hypothesis is that your drug has no effect on pCR, and pCR has no effect on survival... or, on disease-free survival. Another, à la B-27, is that your drug does have an effect on pCR, but pCR doesn't translate into a survival benefit in the treatments. So, in any of those null hypotheses, there's about a savings of a half. This is a random number. It could be as big as 1,800. But most of the time, it's very low.

Under an alternative hypothesis, this is a success -- there's a 25 percent improvement in disease-free survival in the experimental therapy. And, again, there are many ways that that can happen, modulated through pCR or not. The savings is not quite as great. But it's substantial nonetheless.

Advantages: There are fewer patients in the trial. There's no hiatus for setting up phase III. All patients are used in the phase III endpoint. And all patients are used for modeling the relationship between pCR and disease-free survival. "N" is seldom near its maximum. But when it is, it's necessary. There are many trials -- and I shouldn't say many -- but lots of trials that end and you look at the data and you say, "Gee, I don't know the answer." It doesn't mean that you don't know the answer you would like to have known. You don't know whether the drug is effective or not. And you'd like to have continued. It is in those situations that you go to the maximum. And it's in those situations that you get the additional power.

Two reasons for the advantages. One is exploiting the pCR and its potential predictability -- its POTENTIAL; it's a correlate, maybe; maybe it's not related at all. And if it's not, you learn that, and presumably then you have to go to the end. If it is related, you learn that and you get to take advantage of it.

**PRE-OP THERAPY IN BREAST CANCER
38_SESSION 6_7 - BERRY**

Forget this surrogacy business. Ask, "Is it helpful? Is it going to tell me something that's going to allow me to make an earlier decision and to build a more efficient clinical trial?" And the other is this, you know, looking frequently and asking, "Where am I going?"

Further improvements are possible in the neoadjuvant setting. Laura Esserman's I-SPY2 is an example of that. Using biomarkers at baseline, potentially at biopsy when you get tissue, if those are possible. Using imaging -- incredibly important. And we've heard about some of that in this meeting.

And let's get into the current century. Let's start looking at more drugs. I mean, this is, you know... Fifty years ago, we got drugs by squeezing tree roots. Now there are other ways to get drugs. And we ought to take advantage of the drugs that are out there, not have this huge queue. There's false negatives and false positives and false neutrals not looking at them.

And so, using drugs and drug combinations. I mean, as many as ten drugs in the same trial or the same process with various combinations. It's not impossible. We've got to do it. We've got to be building a personalized medicine ability within clinical trials. And possibility of adaptive randomization. So, not only do you stop arms and add arms, but you might discount arms in terms of how much probability you give them because they're not off the shelf yet -- we're still learning about them, but things are not looking as good.

So, that's my outline. I told you about why we have to stop doing adjuvant trials. And maybe neoadjuvant trials are the solution. pCR is a correlate POTENTIALLY; it's certainly not a surrogate. And we've got to model not only pCR and disease-free survival and overall survival, but any other things that come up along the way, like, for example, sera. And we've got to be fine-tuning pCR. Thank you.