going to take advantage of -- I think we -- and so I think the mindset -- given that that's sometimes not going to be the case, the mindset that we have to have the perfect test and a test that's going to last for a decade before we do that pivotal trial or we're going to have to then do another prospective pivotal trial with that test that we evolve and not be able to do bridging studies on archived tissue, I think would just bring a halt to moving to predictive oncology. And so I think we need sort of more realistic standards than that.

DR. DUTCHER: So what is an imperfect biomarker and are there biomarkers that are too sensitive? Dr. Zhou brought up an imperfect biomarker.

DR. PRZYGODZKI: I don't think anybody really knows what is the cutoff and anything to make something positive or negative. That could be misconstrued as an imperfect test.

If you have 95 percent of the tumor cells that show a particular one type of mutation, you're pretty confident that that is really what it is. On the other hand, if you're getting down to the single

cell level, you may actually find that there may be two, three, four different mutations going on. And does that actually mean that there is -- this is truly a mutated tumor? I don't know. I really don't know. I don't think anybody knows.

DR. NETTO: I think that's why it's crucial that whenever you're trying to use the test as a predictive, to use exactly the same methodology, the same cutoff as the pivotal trial or the dredging trial, whatever it is.

So you can't -- that's the problem that you're trying to move into. You cannot use that, yes, mutation should exclude you from treatment and then start just adopting the fancier test that is very sensitive. And you're probably not doing your patient a service because you're not using the same cutoffs and the same standards, the same test.

That's why I think any labeling should come with delineation, what test for that specific label was used? And, of course, that doesn't mean you cannot use others, you cannot study others, but this is what we found, the difference in this setting and in this test.

DR. DUTCHER: Dr. Raghavan?

DR. RAGHAVAN: So I think the problem is we're thinking about this the wrong way, a little bit. And it goes back to the question that I asked Rick Pazdur earlier in the piece. And at the risk of annoying all the FDA people in one sentence, I think you need to go back and reconsider the question that I asked.

What I mean is we're not, around this table today, going to define the ideal test, because it doesn't exist. And while George can talk about the importance of reproducibility of doing a technique the same way in different labs, all of us, including George, who work in labs know that it doesn't work that way. Different labs read the -- the cook reads the cookbook and does things differently, temperatures vary, stuff varies.

So, therefore, the only way to protect the populous at large is to have some mechanism to go back and check. And Rick made the point that the FDA doesn't require storage at the FDA of CT scans, and that's true and appropriate. However, CT scans can be

stored and the better investigator pharma companies keep their CT scans and can always go back and reanalyze or produce them upon demand.

So given that these are your rules and you're setting them, I think if we're going to get into this area, which I think is tremendously important, then we need to have a different mindset about what is it we have.

The FDA has never had difficulty in saying "We want to look at the data and analyze it," and that's appropriate, because FDA statisticians sometimes think differently about numbers from other people and the reasons for that are self-evident.

I'm not suggesting that there should be a gigantic bio repository housed in the Capitol Building, but what I do think is that it would be very reasonable that, as we define new rules, when a biomarker is influencing an outcome as broad as who gets a drug or who doesn't and, particularly, as the test of 2008 will be supplanted probably by the test of 2010, and it may make a difference, I think we need to have some framing reference for doing that.

The other thing that I would say is that the definition of the perfect test will, of course, need to be functional.

So we heard some really elegant data today where the applicants showed us that when a particular test was done, there was a zero response rate in one arm and, I don't know, 30, 50, whatever percent response rate in the other. So the cutoff, when you have zero in one arm and X-plus in another arm, tells you there's a functional difference. It doesn't matter whether the assay is perfect or not.

So one of the constructs that I think we're going to have to have is a functional definition of difference and then we'll have to define what is a difference that's big enough.

Today, we heard -- and everybody in the room understands the difference between progression-free, total survival, but we saw some relatively modest differences in progression-free. Nobody would challenge that they were statistically significant. They probably aren't clinically terribly relevant in this set of studies, but they may be tremendously

relevant by extrapolation to future designs.

So I think we're going to have to think, rather than solving a problem today, we'll need to think much more functionally in terms of the downstream impact of the decisions today.

DR. DUTCHER: Dr. Keegan?

DR. KEEGAN: So I think perhaps we answered your original question a little bit narrowly in saying that FDA doesn't store the samples. But FDA could reach agreement with a sponsor to store the samples themselves under a post-marketing commitment so that samples would always be available for future testing of technology, which is something that we did not as well as we hoped we could have for Herceptin, because I think we didn't know as much about all the important factors for future bridging studies as we found out as we went down to test.

So I think that's sort of the heart of this question is what kinds of things should we think about when we set up these -- if we were to do this, if there was a sense of the committee recommending that there be a post-marketing commitment, that if these two things

are married to each other, that there be a plan for assessment of future technology; what kinds of things should we have in hand looking at the time of approval or, shortly thereafter, approval about a test so that we can do this -- set these commitments up more intelligently?

DR. DUTCHER: Dr. Lyman?

DR. LYMAN: Just to take off on that a moment. And perhaps the agency has thought this through a fair amount, but I think since we cannot see the future and recognizing that virtually any novel or targeted therapy could end up being evaluated based retrospectively on an assay that was not anticipated or not known at the time, that perhaps all -- at least Phase III pivotal licensing type trials of novel targeted agencies should have a mandated prospective tissue acquisition with provision made for storage, not at the FDA, but by the sponsor of the study. And this would just be a recognized required component of any therapy based on some target that could become a functional assay in the future.

DR. DUTCHER: Dr. Wilson?

DR. WILSON: I think that's theoretically a very reasonable idea. But to get back to what has been discussed before, and that is that some of these changes are not stable.

So, for example, if you're looking for a 17p hit in CLL and you're doing it in the relapse setting, but your samples are from the upfront of initial diagnosis, they're not going to reflect what the patients have.

So I think that it's going to really be based on the type of test you're actually going to be doing. I think RAS looks like it's a relatively early event. So it should, in most patients, be present upfront. But other things, such as p53 mutations is very stable on large cell, very unstable in CLL. So it's very much of a moving target depending on what you're looking at.

DR. DUTCHER: One last comment, Dr. Link.

DR. LINK: One comment about the mandate for tissue. I think it's a great idea scientifically, but I'm wondering, especially, because the mandate would have to be for future testing not now specified. I don't think our IRB would approve that, because you

have to have an opt-out and they would view that mandating tissue in order to get on a study is coercive. We've already faced this. And, of course, I'm a pediatrician, where it's even more coercive. But I'm just wondering how we're going to reconcile that with other regulatory agencies.

DR. CURT: I also think that the requirement of having testable tissue in 90 to 95 percent of patients may be a bridge too far. Even in trials where we've required tissue as a prerequisite for coming on study, the actual attrition that occurs in the percentage of patients in whom you can interrogate tissue is actually far less than that.

DR. DUTCHER: So it sounds like a laudable goal, but something that needs a lot of work and we have to deal with the practical aspects of IRBs and pathology departments and dollars and freezers and a lot of stuff.

Yes?

DR. LINK: On the flipside, though, from a pathology standpoint, one needs to hold on to all diagnostic material, at least in blocks, that is, for

ten years and, for pediatrics, 20.

So that tissue is there.

DR. HARRINGTON: Yes, but it may not be consented.

DR. LINK: Oh, that's true. It may not be consented, but the tissue is still there. So the accrual, in and of itself, is potentially possible.

DR. NETTO: I don't think it has to be an all or none phenomenon. I think at least what has been consented and kept be made available for future retesting or looking at subsets of other markers. That's, I think, what we should be talking about rather than forcing no more trials unless you give your tissue. I think that's not going to fly. But if you consented that you gave your tissue, then the company who was introducing the Phase III trial will need to promise that it will make it available and will try to give more information about how the test was run, what were the cutoffs and all that. I think that's important to do it up front.

DR. DUTCHER: Thank you all. I think we're going to move on to topic number one, and this is when

would it be appropriate to limit use of a drug to a subgroup based on retrospective analysis of one or more studies that were not designed to examine this subgroup?

And we're going to ask Dr. D'Agostino and Dr. Lyman to start the discussion and I think we'll probably have a lot of discussion here.

There are a number of points made. I think you can all read the points. Please start.

DR. D'AGOSTINO: The FDA and the sponsors have given us some number of sort of general rules in terms of what should be met and what is essential for doing these types of studies. I think it would be useful for us as a panel to have on record some comments. And what I'd like to do is give a listing of what I think are important issues and turn it over to Dr. Lyman and then maybe a general discussion.

I think that there are probably four -- and you can break these down in different ways. But there are probably four general categories.

Category one, I think, is that these analyses should be hypothesis-driven. Even though they're

retrospective, they should be hypothesis-driven. Exploratory mode is a different thing all together. These are hypothesis-driven with an analysis plan and there should be a validation built in.

So a presentation, say, we are looking at these hypotheses, here's the mechanism for it, here's the way we're going to go about testing it and here's the way we're going to go about validating our results, because we realize this retrospective analysis may still have some exploratory aspects to it.

As far as the efficacy and the safety, what we're looking at, I mean, there is the main trial that we have before us that we're taking subjects from and further analysis and there were efficacy variables. There's the discussion about progression-free versus overall survival and so forth, which I found rather striking in some of these here, but the efficacy is, I think, very much driven by the original study.

The second general point is where do the samples come from, something that we were just talking about. I'm really concerned that no matter how well you plan the study, no matter how well you do your

analysis, no matter what kind of validation you're going to say, if all you have is convenience samples, then you're in real trouble, and some of the presentation of Dr. O'Neill was showing that.

In the studies I'm involved in, quite often, it isn't rigorous, across sites, incentives on how samples are being taken and how they're going to be kept and so forth. And so the idea of having a solid set of samples available, I think, is really going to be driving this. Whether it's 90 percent or what have you, I don't know, but it has to really not be just a simple convenience sample. We have to make sure that the randomization is preserved. There has to be some way of saying that what we have is validated for statistical analysis. We have to make sure that the number is adequate for doing statistical tests and for interpretation.

The third general issue, I think, is the notion of statistical power and multiplicity control. The study should have and should be demonstrated before one starts the analysis that we have enough subjects, we have enough data, enough samples where we can

actually have a good chance of having solid statistical results with multiplicity control built in there. We're going to be looking again at data retrospectively. How are we going to handle the possibility of looking too deeply into it? Again, we have the validation.

I'm co-principal investigator of the Framingham study, and all genetic analysis is basically in terms of we try to control and allow for false positive rates and what have you, but validation on another dataset is the only way that I think the epidemiologic field feels comfortable in. I think we have some of this going on here. But within the study, within the proposed plan, statistical power and multiplicity control.

Then we need consistency, for my fourth sort of general thing, consistency in sensitivity analysis. Do we see consistency as we would anticipate, males, females, young, old, different grades of the disease and what have you? It's not just an overall analysis, but consistency.

Do we find consistency with the interaction

type tests? Those type of tests should be built in. Are the data we have sensitive enough to, in fact, show us that being positive or negative on the biomarker is going to make a difference, that we have an effect modifier, basically, with these biomarkers, and do we have, as I say, consistency across subgroups we would have looked at?

In the sample we're looking at, with the sample we're looking at, regardless of the biomarker, be able to reproduce the original sample results; this idea that we aren't dealing with a unique sample, we're able to show that what we get from this smaller sample, or the sample that we're doing this retrospective analysis, can reproduce anything that was in the original study.

I think if we have these -- as I say, these are four general things and I think they just reflect what was mentioned earlier. But I think if we have these things met, these categories met, and we say that these are the sort of general categories one has to look for in a study, you could, in fact, put an analysis together that could, in fact, be believable

and not left just to sort of our feeling good about it.

DR. DUTCHER: Thank you.

Dr. Lyman?

DR. LYMAN: I certainly agree with all of that. I'd just take one step back.

I personally feel that if the original endpoint was not reached in the prospective trial, the desire to go back and look at subgroups retrospectively just doesn't really hold and I think we should require, ideally, two prospective studies for patients who are stratified a priori based on subgroups of the assay or the treatment was limited to a specific subgroup of the assay.

In the situation here, where there may have been the -- the outcomes were reached, the primary outcome was reached, then I think we have discussed -- and Dr. D'Agostino has listed many of the issues that need to be addressed.

It's extremely important with these type of analyses that they be adequately powered within the subgroups. Particularly, if the marker may be both prognostic and predictive, there has to be sufficient

power in the better prognosis group with fewer events; that if there is no demonstrated treatment effect, we can rule that out with some level of confidence that's high, and that's often hard to do.

There has to be formal testing for drug-biomarker interaction. That's been discussed previously. There needs to be appropriate adjustment for all other prognostic and predictive factors that are known. Obviously, we don't necessarily know all of them.

While, again, I recognize it's a high mark, we've seen that achieving samples, retrieving samples, in 90 percent or greater, the acquisition rate can be reached.

I think we have to set the bar high if we're going to do this type of retrospective look at prospective data. I think to look at data where only half the samples, or in some cases, less than half the samples were obtained, leaves us open to all sorts of potential confounding that cannot be fully resolved retrospectively.

As Ralph mentioned, obviously, the analysis

should be blinded for the biomarker and the analysis should be pre-specified before those markers are on.

Again, we all pretty much agree what the major quality issues here are. I think they just need to be transparent, known to the sponsors and trialists in advance, and we should insist on those types of criteria.

DR. DUTCHER: Thank you.

Dr. Harrington?

DR. HARRINGTON: Thank you. I just wanted to make a distinction here that I think is well understood, but certainly I see it on both sides.

There's a difference, obviously, between regulatory approval and the march of science here. And so I just want to be sure that we don't send a message as a committee that you can't learn anything from a trial that didn't meet its primary endpoint. It's important to try to learn from those trials. It may need future prospective trials for regulatory approval, but we certainly don't want to imply that these exploratory analyses aren't useful, as risky as they are.

The other thing I want to say -- and I might come back to this when I talk about the next question. I benefitted from a very useful and mildly illegal discussion with my colleague, Professor Simon, Dr. Simon, over lunch. We weren't supposed to talk, but he set me straight on this issue about the ascertainment.

I think we also need to be a little bit careful about that, because, in fact, if you had full ascertainment and then you did a study on ten percent, a random sample of the ten percent of the blocks, you would have a perfectly unbiased sample to look at the effect of a marker within groups. So the issue in this ascertainment isn't so much how many you have, it's how you got them. And I think that acknowledging the fact that you can't get them all, it's really important to understand how you got the ones that you did.

There are certainly situations where these convenience samples can lead to bias. So maybe the easiest one to understand is let's suppose that in the mutated group, there's no treatment effect, and let's suppose that in the wild-type group, there is a

treatment effect, but it varies across treating institutions for various reasons that are correlated to patients, and then you only get the blocks in the institutions where they saw the large treatment effects. So you'll get a biased estimate then of the effect.

So I'll say more when we go on to the next question about the value of full ascertainment, which I think is important, but I think throwing out a number, 90 percent, and then treating that as the reason, as the only way to measure whether ascertainment is perfect or not, is a bit shortsighted.

DR. D'AGOSTINO: Can I respond?

When I was making my little presentation, I said I don't know what the percent should be, but it's the convenience.

But I do agree that if we leave it loose, that it's not a convenient sample, you can compensate by saying we're going to have enough power. So I think you can get at it by saying your sample is going to be adequately powered.

The other thing in terms of the negative

study, I was bothered this morning, and I raised a question, I think it was probably the first question, but it is possible, if these genetic factors or biomarkers are so important, that you have a positive effect if you're positive on the genetic marker and you have a flat effect if you're negative.

In the overall sample, it could be that you're sort of washing things away. So I don't know how I could do the mathematics out and say that it's possible to have a negative study where there's a really powerful subset, but it's conceivable, and I don't know how to handle that in terms of what we're talking about here. Would that negative produce so much noise or would the overall be able to show it? And I don't know. I don't know what the answer is to that.

DR. DUTCHER: Dr. Simon?

DR. SIMON: Well, the couple of papers that were alluded to this morning that I published sort of got into those calculations, and that's actually the typical situation.

Other people with Herceptin have shown that

had those trials taken all comers and not been limited to patients whose tumors over-expressed HER2, they would have needed -- because the incidence of positivity was only 25 percent, those trials would almost certainly have been non-significant. And to get them significant, to make up for the dilution effect, they would have had to have included thousands and thousands of patients.

DR. HARRINGTON: You were always on my mind when I was making my comment.

DR. SIMON: But I wanted to touch on -- and so I also believe -- we've had this sort of conventional wisdom, never trust subset analysis unless the overall results are positive, and that has sort of protected us against data dredging.

But what we're talking here, we don't need -- that is actually sort of an irrational rule of thumb now in terms of what we're really talking about and we don't need that to protect us against data dredging.

So we need to distinguish data dredging from the kind of KRAS situation, as an example, we were

seeing today. But if we continue to use this rule of thumb, never look at a trial unless it's met its over -- it's significant for its overall, that leads to clearly erroneous conclusions. And so that rule of thumb really needs to be sort of given up and we need to independently make sure we're not talking about a data dredging situation.

The other thing is -- I guess the big thing is we need to distinguish -- for example, the kind of prospective/retrospective design and the conditions for doing it that were presented this morning by the FDA I think are very useful, and we need to not lump those kinds of analyses together with the sort of typical data dredging analyses.

In other words, the key things are the kinds of things that Dr. D'Agostino was talking about. It needs to be a focused analysis. It needs to have enough patients, both in the test positive and the test negative subsets, to be interpretable, and you have to have a test that is analytically validated on archived tissue.

But I can conceive of situations where you

could do an analysis -- even though the trial was big enough and the proportion positivities were appropriate and you had arranged for archived tissue, and you could actually do, to me, just as believable analysis if information arose during the course of the trial from external sources as if you had set it up from the start that way.

That may not be the typical situation, but I don't think because it wasn't done completely prospectively that that precludes being able to -- if other things are right -- being able to reach reliable conclusions.

It was alluded to this morning that this term kept popping up, stratified, randomized stratified by the prospective -- by the predictive biomarker, meaning that the -- if, by stratified, we mean that the randomization is balanced by the predictive biomarker, that is not, to me, a viable objection.

That is not, to me, an essential. You can do a perfectly valid randomization test without prospective stratification and all of the prospective stratification -- if you know the predictive biomarker

in advance, then prospective stratification is valuable because it assures that you will have tissue and assays for all of the patients who go into the trial. But it doesn't really do anything to improve the validity of the analysis, and it doesn't actually improve the -- all it improves is the balance between the number allocated to treatment versus the number allocated to control for, say, the test positive patients. It doesn't improve the balance of those with regard to unknown covariates.

So there's, I think, a lot of confusion about the supposed benefits of prospective stratification, at least as it applies to sort of providing a basis for inference. I think key issues are sample size, multiplicity control, having a focused analysis, and those types of things.

DR. DUTCHER: Thank you.

Dr. Link?

DR. LINK: First of all, I'm glad to hear that Dr. Simon is supporting the rationality, because now I can justify why I bought a lotto ticket.

But actually, there's a good example of

retrospective. Look at this lung cancer trial with EGFR inhibitors, where a very small subset of patients benefitted hugely in an otherwise negative trial. And I think that maybe Dr. Harrington's comment that the difference between discovering a potential marker versus regulatory approval of that may be relevant. In other words, there's good examples of being able to find markers retrospectively. So I can't imagine that we would want to eliminate that possibility, but maybe the FDA wouldn't accept that kind of approach to put it on the label.

DR. DUTCHER: Dr. Raghavan?

DR. RAGHAVAN: I think we also want to remember the comments from the first of the patient advocates, which was essentially a plea for common sense.

And while I agree totally with what Richard Simon said, as I have over the years, the reality is that we want to be careful that we don't box Dr. Pazdur and Dr. Keegan into a little corner where, with our information, we set a bar that's so high from our advice that they can't make sensible decisions.

One of the attractive features about ODAC is it doesn't have lawyers on it and so we can actually think about patient welfare. Sadly, worldwide, our treatments aren't that great in many domains, and as the patient advocate said, if you have a technology that might restrict usage inappropriately, that's probably a good thing.

I think one of the things I've felt has been lost a little bit today, because it's probably one of the very first times I've seen it at the FDA, is two companies have come here to try to create a situation where they sell less product. That seems like kind of an important thing. And so, therefore, perhaps the way we need to think about this is in terms of, yes, we need to set rigor, we need to have good assays, we need to have well powered studies. But we might create a fudge factor that would let Dr. Pazdur, et al, look at the numbers of sets of data, the overall numbers.

If you think about our clinical trial domain, we've come up with a crooked trick of meta analysis that allows us sometimes to glean information from rather poorly executed studies, where the numbers are

small. That's not a replacement for a very well designed randomized trial.

But the point I'm making is I think if we set rules that have common sense in them and allow the FDA some discretion to look at what was the intent of the study -- as Mike Link said, I think, were you able to glean a useful quantum of reproducible information, even though the study wasn't designed to do it. And as I've been hearing the discussion, I've been a little uneasy that we're starting to raise the bar with a lot of clever terms that will actually stop common sense from being implemented, and that would be a shame.

DR. DUTCHER: Dr. Zhou?

DR. ZHOU: I have two comments. One is actually related to the ascertainment, because that just makes me think that if we think about missing data, the importance about missing data is not how many people we're missing, actually, it's about missing information rate.

So we could actually borrow that kind of thinking into the ascertainment areas, instead of a proportion of people don't have a biomarker and maybe

to say what is the missing information due to the missing biomarker. Maybe that's more important than to say what are the number of people missing.

The second comment I wanted to make is related to the subgroup analysis and overall group analysis.

So actually, sometimes the subgroup analysis can help us to better analyze the primary analysis. Let's suppose the treatment actually does depend on which biomarker value you have. Then you cannot ignore that when you do an overall analysis, just do two-sample comparisons. So you have to take the effect -- the treatment effect is actually different depending on which biomarker you are using.

So in other words, I think performing the subgroup analysis sometimes actually can help us to inform us how to analyze primary data analysis, also. So that's an advantage of performing subgroup analysis.

DR. WILSON: I wanted to say that I was very happy to hear Richard's comments on it not necessarily being needed to have prospective biomarker stratification for a practical, as well as a scientific

reason.

Number one, it is often impractical to do these tests before people come on study. Number one.

Number two, you think, going into a prospective trial, particularly in an early drug, that you may know what the proper biomarker is. If, in fact, you do do stratification based on that biomarker and then later on find out that it's really something else, you have severely biased your groupings; whereas if you start from the very start blinded to the biomarker, I think it makes the trial much more amenable to future looks with other markers.

DR. DUTCHER: Ms. DeLuca?

MS. DELUCA: Thank you. When the emerging science meets the road, it's usually on my body and bodies of probably people that are in the room, and it's our lives and our bodies that come to you from the bedside. We don't always stay in the bed. We also walk around. We breathe. We'd like to go to holiday parties, see grandchildren's birthdays and that sort of thing.

We're talking about a lot of numbers. I'd

like you to think, one, on the numbers of people who have already made their gift to you of the tissue samples. Don't waste them. Please keep them. And then take a look at how many do we need for the future. Let's take a look at that number. How many people? We're talking about metastatic colorectal cancer, we're talking about more people in stage three and stage four than we are in stage one and two and zero. Also, no matter which progression you're talking about, they die.

So that's our bottom line. So, please, use the gift that we have, take it. The statistics are very important, vitally important to us. KRAS is a subject that's vitally important to us.

You can Google, go to Yahoo, any of the sites, put in colon cancer, put in metastatic colon cancer. Out of your top ten picks, you're going to find eight of them are going to contain KRAS. So this is a subject that's very important.

So my question is more, how many are we going to need? We have thousands, looking at the trials that have been presented today, between the trials, the four

trials, the six trials, if we bring in European trials, thousands of people who are already represented here. How many more do we want? Do we want thousands more or are we talking one or 2,000? I think that's something that should be sort of determined upfront. Then the other questions that have been brought to fore would really make much more sense and file in.

Thank you.

Oh, I'd like to say one more thing. I was really pleased when Dr. Little got up, because I didn't know. I thought DxS was a type of assay, but I didn't know that it was a U.K. biomedical company. I was thrilled to know that.

Thank you.

DR. LITTLE: Glad to be of assistance.

DR. DUTCHER: Are there other comments on topic one? We should move on.

Topic two is going to be discussed by Drs. Harrington and Richardson.

And this is when would a prospective study -- well, do we need to -- wait a minute.

Should we summarize topic one? Did we get to

a summary? Let's go on.

When would a prospective study design for the purpose of examining treatment effects on a pre-specified subgroup be needed to establish treatment effects in this group?

So I guess, and otherwise, when is the retrospective data not strong enough?

Dr. Harrington?

DR. HARRINGTON: Thank you. Let me touch on a few points.

First of all, I'm going to agree with Rick Simon, in principle, that if there are two very good retrospective studies that meet the criteria that Dr. D'Agostino and others have pointed out, then I'm not sure that we do need a prospective study.

So I'm going to talk about the situations I think that weaken that evidence in the two very good prospective/retrospective studies that would point to the need for another trial.

So the first is this ascertainment issue, but it's not the ascertainment issue of the specimens. I'll talk about that in a second. It's the FDA's

ability to ascertain all the studies. So I think that if there are two very good ones, that doesn't mean that there aren't ten out there that showed that the marker wasn't informative. And so I think we would need to know that the FDA was able to capture most of the available data that was done in well controlled trials and that it was consistent.

The ascertainment process of the tissues we've talked about a lot. So I'll just say it's important to understand that process and if there's any loss there, to understand where the loss is coming from. And if we can't understand that, then I think that would point to the need for another trial.

If the important endpoints of the prospective/retrospective trials were contradictory, progression-free survival, overall survival response, if they didn't seem to all run in the same direction, not necessarily all significant, but all running in the same direction, then I think a prospective trial could be much more important.

The lack of a pre-specified analytic plan, that's been discussed. I won't say anymore about that.

That's, I think, a sine qua non. You have to have that.

I think that we need to be very careful, I think, here about the patients who appear not to benefit because of their biomarker status and to make sure that we're not experiencing a Type II error there, an error of false negative. And so there, I think, it may well be that there are instances where the subsequent prospective randomized trial may need to be in the biomarker negative, let's call them negative group, if there were intriguing trends in that group, but non-significant. So we should just be careful, I think, not to leave them behind.

I think there are other situations that have been mentioned, as well, situations primarily where the science has changed substantially since the original trials were done, either in the way that the marker is being done, in the way that the treatments are being given, and our, perhaps, change of heart about an intermediate endpoint, like progression-free survival, that was used in the earlier trial and now we begin to wonder if that really is the best way to measure

things.

So I think that in summary, my default would be if we have two really good of these retrospective/prospective trials that meet all the conditions that have been specified, and they are the universe or nearly so, then I don't think we need a prospective trial. But absent any of those conditions, that's when we need it.

DR. DUTCHER: Thank you.

Dr. Richardson?

DR. RICHARDSON: I have just a couple of comments. One is these obviously are so complex, when we're looking at these various biological mechanisms, that I think we need to spend more time in trying to make sure that we have the appropriate markers that we're studying.

Obviously, you could argue that in the data that were presented earlier, if fewer than 20 percent of the patients actually have an objective response in the more favorable group, the situation is very complex and requires further studies with trials looking at the proper markers.

Without spending a lot of time trying to define the word "was," I'd like to look at the word "when," because with regard to conduct of these studies, looking from the perspective of a clinician, I was struck by the repeated assertion this morning that a randomized study of some of these drugs in wild-type KRAS colon cancer patients can't be done. We listened to Dr. O'Neill's very, I think, elegant analysis of these data and one would wonder whether that conclusion that these studies can't be done is true.

I think this gets back to a couple of issues. One certainly is one of balancing risks between two groups of patients. How do we go about doing that when the information that is out there, as Jo-Ellen mentioned, on Google and out on the Internet is so directed in a particular orientation?

How do we deal with the prejudices and biases of physicians? Because once these issues make it to the speakers' bureaus, those also make it difficult to design these studies and execute them, in particular, if accrual falls off because of various changes in judgment on this.

So I think it becomes imperative to pursue these kinds of studies in a very timely fashion and I think everybody, whether we're dealing with cooperative groups, whether we're dealing with the sponsors, but everybody needs to show some judgment and restraint on this.

Finally, there's one little issue that's always bothered me about these kinds of biomarker studies and that is that there really isn't anything out there indicating that one actually achieves, at a cellular level, what you think you're hoping to achieve. I don't know of any data that have been presented indicating that, yes, the drug is there and it's doing what it's supposed to do. I think as technology evolves over time, that would be something that should be a real goal for this.

DR. DUTCHER: Thank you.

Comments? Dr. Curt?

DR. CURT: I think this issue of ascertainment of all clinical trials is an important issue, as well. There's been recent publications by Scott Ramsey and others on publication bias, where

negative clinical trials just don't make it to the peer reviewed literature. And I think it would be good to have a mechanism that such trials that are well done, but don't meet their endpoints, are published, are searchable in Medline and Medlar and just don't get posted on a Website somewhere.

DR. DUTCHER: Dr. Simon?

DR. SIMON: Just to agree with Dr. Harrington's statements about the two, well-done retrospective/prospective studies and that those be -- or that the ones that are done sort of are consistent and that there be at least two good retrospective/prospective studies.

In terms of the definition of what makes for the good retrospective/prospective study, I think the only thing I would differ with what the FDA presented this morning was their very high definition of required ascertainment, that 90 percent or so of the specimens be ascertained.

I think there actually is a bit of a confusion there. I think the ascertainment percentage influences generalizability of results, but it doesn't

actually introduce bias, as long as the ascertainment is not differential by the treatment groups.

So of the cases for which you know the biomarker result, if those are properly randomized cases and if treatment has not determined ascertainment, then you wind up with an unbiased estimate and a perfectly valid test of treatment effect in, say, the test positive patients for whom you have tumors assayed.

The only issue is are they representative of the entire group of patients in that trial, which is similar to the situation you have even in a fully prospective trial, where you sort of never really know whether the patients in your trial are representative of the populations of patients outside of your trial.

So it's a little bit more difficult when you're not even sure whether they're representative within your trial, and that's why Dr. Harrington, I think properly, emphasized you really need to get into detail of who you have tissue on and who you don't. And if your ascertainment percentage is very high, then you don't have that worry. But qualitatively, it's not

really any different than the issue of generalizability in a fully prospective trial.

DR. DUTCHER: Dr. O'Neill?

DR. O'NEILL: Yes. I'd just like to follow up with Rich on that in terms of caveats he would put on the diagnostics that you would assure yourself that you are in the sort of comparing likes with likes.

The whole issue is that that's unknown, usually, and, empirically, all you can do is say, "Well, I don't have differential ascertainment treated and control group." Well, how do you know that? You have to have some access maybe to perhaps the universe source to know that the sample that you do have is relatively representative.

So do you have any advice on sort of the diagnostics that you would look at to assure yourself that you don't have a biased sample?

DR. SIMON: I think the only advice would be you want to know all the details about the cases, whether there's institution variability in treatment assignments and all of the covariates and as much as possible about the issues of who you have the

ascertainment on and who you don't, so that you can try to assure that there's not a treatment difference on ascertainment and, also, to try to understand what potential issues might be there in terms of generalizability.

DR. ZHOU: Actually, I think that in the literature, they do have some -- this is actually a related issue similar to the meta analysis issue about whether the study you select actually represents the whole population.

The Cochrane, in the clinical trial, also, in diagnostic medicine, they do have a guideline that has published to say what are you looking for; for example, sample size and study population and how to deal with missing data. And then they give the score for each study as published and then they come up with the quality of the study score and then try to judge whether -- suppose that you have six studies, for example, you show us in the morning, and then they tell us to say what's the quality of each study, based on the information available.

So they are looking at factors, not just

ascertainment rate. They also look at other factors they think are important. So we should apply that kind of criteria to the biomarker study, also.

DR. D'AGOSTINO: We all have different experiences and so forth. But quite often, the studies I'm involved in, there's a hard and fast protocol that one follows, but then there's sort of secondary things you're doing and there are rules, but not necessarily adhered to. And I think that's what -- might be getting to, that there isn't uniform ascertainment, and sometimes it's from one center to another.

So I'm not so sure the Cochrane type of rules, but I think more the Richard Simon type of field might be what you have to apply. There's not going to be a hard and fast that you can actually pull an answer.

DR. SIMON: Ascertainment issues can be much more problematic in situations where you're talking about follow-up data and lack of --

DR. D'AGOSTINO: Some of these things are part of the follow-up in terms of samples and things that you're getting, that you aren't necessarily

getting them all at baseline; that you decide later on to get certain variables and so forth on the subjects and you only get them when they come. And then other centers are doing sub-studies where you get certain information and others aren't getting it, that type of mix.

DR. SIMON: I'm saying those are very complicated situations. Here, what I think we're talking about is just a baseline sample. And so I'm saying the issues and the opportunities for bias, I think, are much less.

DR. D'AGOSTINO: Well, prospectively, but I think we're talking here about there may be loads of studies that you might have the ability to do some biomarkers on whatever data is available from those studies, and it's not that you're now looking at prospective collection, but it's whatever you got in the previous running of the study.

DR. SIMON: But these are not surrogate endpoint kind of biomarkers. These are predictive biomarkers that have to be based on samples taken essentially before the patient was randomized.

DR. D'AGOSTINO: Then why don't we have 100 percent?

DR. SIMON: Because -- well, there's lots of reasons why. It's difficult to get the tissues sometimes from the pathology departments, things like that.

DR. ZHOU: Assuming all these studies are actually is published. Right? So you should be able to find all that information for the Cochrane.

DR. DUTCHER: Dr. Funkhouser?

DR. FUNKHOUSER: My perspective from the pathology side is that patients who consent to participate in these trials frequently have extra tissue sampled at the time of resection.

So for example, on a colectomy done for primary colorectal carcinoma, they'll sample three of the primary for staging, but then separate tissue will be passed off either to the tissue bank or to the nurse coordinating this particular trial, and then sent to central pathology for bio banking and DNA extraction and so forth. So I don't see why 100 percent ascertainment isn't possible in a trial that

anticipates doing these sort of prospective/retrospective analyses.

DR. DUTCHER: Well, I think that's true in academic centers, where there's somebody that will put the tissue in the tissue bank. But if you're doing a Phase III trial where you're including community oncologists who are treating the patients, they may have their tissue in the pathology department in a community hospital.

DR. FUNKHOUSER: My perspective is that if the hospital commits to participate in the trial, then the pathology department should cooperate with that, and you should --

DR. DUTCHER: I agree, in the best of all possible worlds.

DR. FUNKHOUSER: She cut me off. I can't believe that.

You should asymptotically approach 100 percent ascertainment for these trials.

DR. DUTCHER: A laudable goal.

Any other comments on topic two? I just have one question that would come through here.

Is there a fraction of missing biomarker data in a retrospective study that would make you say you must do a prospective trial?

Dr. Harrington?

DR. HARRINGTON: No. I mean, if it gets down to near zero, obviously, there's not enough information to analyze. If it's at near 100 percent, it's perfect. But if it is 50 percent, but it was a 50 percent true random sampling or through some process that's essentially random, then that 50 percent may be useful and, as Rich Simon said, would produce an unbiased estimate.

So it isn't only about the size of that group, it's how they got there and whether there is something that might be correlated with treatment or outcome that's hidden.

DR. DUTCHER: So I think that's important for the FDA to hear, and how you would set those guidelines, because they're looking for guidelines in terms of how you look at these data.

DR. HARRINGTON: Statisticians love to qualify. So I'll just qualify that by saying the more

tissue you have, the more ability you have to cross-validate your analysis or to validate it. And so more tissue always results in a stronger analysis.

DR. DUTCHER: Dr. Funkhouser?

DR. FUNKHOUSER: Thank you. Just to address topic two, it's my understanding that a prospective study would be sold to a potential patient to be accrued on that study if and only if you didn't know which therapy was best. And it seems to me that the data that we've looked at this morning shows no evidence of even a partial response in patients that are RAS mutant.

Is that correct?

So I don't know how you would accrue patients on a trial if you already have evidence that there is no potential benefit and yet a 25 percent probability of side effects for treatment, given RAS mutant status.

DR. DUTCHER: Dr. Raghavan?

DR. RAGHAVAN: I think that, not as a statistician, but as someone who spends a lot of time with clinical trial statistics, the further away from 100 percent you get in terms of ascertainment, the more

red flags go up. And while I accept Dave Harrington's point totally, because, again, I was the one who said use common sense, I do think the FDA needs to have some rule that says the further you away you are from 100 percent, more red flags go up.

What was cool about the data today -- they were pretty good data. There was high ascertainment, with the exception of one trial. And what troubles me is that when this sort of a presentation comes up, it's always the A team that's presenting its data.

Downstream, Rick and the gang are going to be looking at retrospective/prospective studies with 25 percent sampling and outstanding data for a quarter of the patients involved and some nonsense about why the other 75 percent didn't get ascertained; they were being investigated in Benghazi or the plane crashed or whatever.

So again, I was the one that said let's not put in too many rules, but I would also say let's keep common sense there. The more people that aren't ascertained, the more risk there is that there's something crooked going on.

DR. DUTCHER: Dr. Simon?

DR. SIMON: The only thing I would also say is that, to me, what was compelling about the data presented this morning was the replication of it in different studies rather than the particular percent ascertainment in any one of them.

DR. DUTCHER: Dr. Pazdur?

DR. PAZDUR: Concerning this whole area of ascertainment, there were a lot of caveats put on the comments; if you believe that randomization is preserved, if you believe that there's no bias here.

How do you really determine that in the real world, that there's no bias that exists in the attainment of these samples?

We all know that there are many things that go into obtaining informed consent in acquisition of samples. Good performance status patients may be having more readily acceptable samples than poor performance. Various countries having more acceptable behavior in obtaining these samples.

So how do you really deal with those ifs? And I think that's the major issue here when one really

puts all these caveats in their answers.

In the real world, how do you deal with that?

DR. DUTCHER: Dr. Harrington?

DR. HARRINGTON: So I think what you're hearing from at least some members on the panel is that the simplest rule to set is get all the tissues and that eliminates the possibility that something is selectively missing. But we're saying that that may be practically impossible and may eliminate the possibility of lots of good science and lots of good results when the loss of that tissue did not disturb the randomization.

Now, the question you're asking, I can't answer. Rich tried to answer it before and I think he struggled with it a little bit; how do you look at the available ascertained tissue and decide that you haven't introduced a lack of generalizability, as Rich likes to say?

So the way I do that in the clinical trial is very labor intensive. I look at every case that's been eliminated -- hopefully, there are only a relatively few number of them -- and try to understand, in the

hidden meaning there, whether there might be some selective effect there that is hidden even from the investigators.

DR. DUTCHER: I would just like to say one thing that Ms. DeLuca mentioned that we sort of said "thank you" is the patients. This needs to get on the Website along with KRAS, that the tissue is very important, because we spend a lot of time with informed consent and explaining why we need tissue and what it's going to mean and it might not help you, but it might help somebody else, and some people buy it and some people don't buy it.

But I really think if the patient advocacy community understands, which many of them do, the importance of this prospective effort in community hospitals and in private practices, in addition to academic centers, that that may be a way to meet Dr. Funkhouser's goal, because he says it should happen. He says that there's no way it shouldn't happen.

DR. FUNKHOUSER: There's one exception and that is if the primary is very small. Think

sub-clinical breast carcinoma. Because radiology is so good now, we're typically seeing breast cancers that have to be triangulated with needles to even know where they are.

So that's an exception. We're not going to give away any tissue to a tissue bank or to a trial that could conceivably interfere with our ability to accurately diagnose and stage the patient. But big primaries, colorectal, lung, there's no reason on earth that we shouldn't be able to collect fresh tissue for frozen specimens and paraffin blocks for trials.

DR. DUTCHER: Dr. Wilson?

DR. WILSON: I just wanted to make two comments just in terms of how do you handle the problems with ascertainment and the fact that you may have a biased sample. I think the way you do it is what Rich said and what was demonstrated here, and that is that you have multiple studies showing the same thing. I think that addresses that very well. How many you need, I think Dr. Harrington said two well performed studies would be adequate. I think the devil is in the details, but I think you need multiple

studies showing the same thing.

The second point I wanted to make is in terms of tissues, and I think it is very important to put the need for the tissues on Websites so that patient advocates realize that.

But as somebody who does a lot of biomarker work and who is now doing a prospective study, I'll tell you, the biggest block isn't the patients. It is the treating doctors. That is what we really have to deal with. And again and again, I keep hearing it's the patients. But the enemy is us, and that's where the roadblock is, I believe, besides the obvious issues with funding and storage, et cetera.

DR. DUTCHER: Ms. DeLuca?

MS. DELUCA: Thank you. Before we get off of this topic, the patients, my heart goes out to the patients who have been told, or even not told, that they have the KRAS mutation. Some people are just sort of silently let off the study or a nurse will kindly say, "I'm sorry, but you didn't make it."

It would be really good to keep them for another trial. It would be really good to keep that

tissue by just paying a little attention to telling them, because in many cases, they have had to give up their insurance to be on these trials. Almost every protocol that I've been looking at in my center is now saying yes, but if you've been tested for KRAS, now your Blue Cross isn't paying. So that's something we should remember for them.

Thank you.

MS. MASON: Well, I'll belabor this tissue issue a little bit longer. My organization has a bio repository of tissue and it's patient-driven. And one of the problems we've had with patients being able to access their tissues is that embedded in their consent, when they were treated in a hospital maybe, that they give up rights to any of that tissue or to be able to direct it someplace.

So there are lots of legalities around that and it can vary from state to state. But I think patients are becoming much more savvy about it, with the help of organizations. There's one group that's done some great work with two brochures and they talk about tissue is the issue.

DR. DUTCHER: Dr. Zhou?

DR. ZHOU: I think a study to check the ascertainment bias is a good idea. However, I think we also have some guidelines about what to do when you have only one study. And there are several suggestions I can think about, which the sponsors this morning actually already have done that.

So first, you can check whether, at the baseline, the covariates are different between the people who have the biomarker results and those regional data. And also, you can also look at are there any major confounders which would maybe different between those two groups.

The second is you can analyze data two ways. One is so-called completed analysis, which most people have done here, which is you only use the data who has the biomarker information. So that's called complete data analysis. And then compare with the other analysis which uses all the data, which one of the sponsors, I think, talked about in the morning. But if they do multiple comparison, that's more the better.

So if you have a different way to analyze, if

all the results are consistent, then you have some assurance that you may not have bias in your data.

So think the FDA should maybe think about providing some guidance about how to do that for one dataset, not multiple datasets, because sometimes you may not have multiple datasets.

DR. DUTCHER: Dr. Keegan?

DR. KEEGAN: With regards to the consistency issue, since the effect appears to be consistent across PFS, but not across OS, to what extent is that a consideration? Because I know that, at one point, there was a statement that the endpoints in the trial should be consistent with each other.

DR. DUTCHER: Dr. Raghavan?

DR. RAGHAVAN: I'm a big believer in OS for many things, but in this particular context, there are so many confounding variables downstream.

So I wasn't too bothered by the fact that there was a discordance between progression-free survival and overall, because the companies didn't attempt to tell us what happened after the first round, and so they could easily have been non-random spread.

SSo I wouldn't be too disturbed.

I think the key is to have comparison of apples with apples. So the agency can say, "We want to look at PFS and OS," and then look at data downstream from that. But I wasn't compelled that there was a problem with this morning's data at all.

DR. HARRINGTON: The other issue is that the replication might trump that.

DR. RAGHAVAN: Correct.

DR. DUTCHER: Dr. Grem?

DR. GREM: I think to answer Dr. Keegan, part of the difference in outcomes was how the studies were designed. So for the panitumumab, they specifically planned to allow crossover to the patients. So a lot of people would say, "Well, that's nice," because if you get randomized at best supportive care, you still have a chance to get the study drug later. But by doing that, you have to throw out -- you figure that progression-free survival has got to be the primary endpoint, whereas the other study that was done in Canada specifically was targeted to demonstrate a survival advantage. So, therefore, those patients did

not have access to cetuximab, except for in the study, if they happened to get randomized to get the cetuximab; if they were on best supportive care and progressed, that was it, they did not have access to cetuximab.

So I think that it's one thing to say, "Well, gee, we're not seeing consistency," but I think you have to look at the trial design and what the endpoints were. And you can make arguments in favor of both designs, both drugs were approved, but the intent of the studies was quite different. And so I think it's okay that one just focused on progression-free and one focused on overall survival.

DR. DUTCHER: Dr. D'Agostino?

DR. D'AGOSTINO: When I was responding to, one, I threw out that in terms of the efficacy variable, we have to look at the what the efficacy variable is in the original study and how it looks, and I was actually concerned about the progression-free survival and the overall survival. And it may be the explanations that we're hearing now, but I walked away from those studies and some of the analysis feeling

very uncomfortable that there wasn't consistency or didn't appear to be consistency. And we were sort of focusing on a different set of questions. But I don't think you can walk away from that question easily.

DR. DUTCHER: Dr. Wilson?

DR. WILSON: I think this brings up the bigger question of whether or not response rate and PFS, which are, indeed, surrogate for benefit, really have a place in disease settings like this, and I think that's really the bottom line.

I would agree with Jean that if you are, in fact, targeting different endpoints, there are design issues that could preclude your seeing other endpoints, such as Jean pointed out.

But I think in the absence of a quality of life benefit with PFS, I, myself, for non-curative diseases, think we need to be moving more and more away from these surrogates and more toward the bottom line, which is are you living longer and are you living longer with better quality of life.

DR. DUTCHER: Dr. Simon?

DR. SIMON: Well, I'm not going to try to

talk about what should be the correct endpoint for approval for a given setting of colorectal cancer, but I would say based on the data given this morning, the consistency of the PFS data and the response rate data, to me, was just sort of overwhelming. I don't know what else you could ask for. It was sort of a slam-dunk, as far as I was concerned.

You can always find something to say, "Well, gee, I can worry about this or I can worry about that," but I just found the data -- and I think you use PFS, because, presumably, I guess that's the most sensitive sort of endpoint and there was not even a claim of a -- when you're talking about third line treatment of colorectal cancer with a single agent, with a crossover, weeks after the patient progresses, do you really expect to see an effect on survival?

DR. DUTCHER: Okay. Well, we're doing very well. Let's move on to topic number three. This is going to be discussed by Drs. Harrington and Raghavan, and it is discuss the properties of clinical trials originally designed for non-selected populations that would make such studies unsuitable for demonstrating

efficacy in a biomarker subgroup.

Discuss in your answer potential problems associated with the failure to perform stratified randomization based on biomarker status, failure to pre-specify statistical adjustments for multiplicity and incomplete ascertainment of biomarker convenience sampling.

We hit on some of this already.

Dr. Harrington?

DR. HARRINGTON: We did. Thanks. I'm wondering if, actually, Rick has given us the same question sort of disguised to see if when replicated, we're consistent with our answers from question to question.

DR. DUTCHER: You've got it.

DR. HARRINGTON: I do think, actually, that we have answered this one, most of it. So I think, just to get started on this -- well, the only thing I will add is that the critical pathway initiative is terrific and it was well laid out. And I think what I'm hearing in the committee is that -- and perhaps this is why we're having the meeting -- is that the

agency will need to develop a set of useful working guidelines for situations where the critical pathway just didn't apply, because either science got ahead of the marker development or the marker development got ahead of the clinical trials or something.

All these topics, I think, have been discussed. The one that jumps out to me is the failure to pre-specify statistical adjustments for multiplicity, which is always important.

I think what Rich Simon was emphasizing before is that we shouldn't treat these as data dredging exercises, that you simply count up the number of tests you're going to do and then divide your P value by that number, but you should really force people to say there's a pretty good biological story here that's behind the scenes that generates a hypothesis or a question or an analysis, and that's the primary one that we're going to get at. And I think the sponsors made the point this morning that when they went back retrospectively, they looked at KRAS. They didn't look at a panel of markers.

So I think that's really important and it

isn't necessarily done through some sort of statistical trick like a Bonferroni approximation; it's done through limiting to a focus question, what you want to look at in these studies.

We've talked about the ascertainment of biomarker and I think Dr. Simon made a perfectly compelling case about the stratification. It's just very hard to do that. It may not even be useful in these settings.

DR. DUTCHER: Dr. Raghavan?

DR. RAGHAVAN: This is an opportunity to be self-repetitive and I apologize. But I think what perhaps what Rick was looking for was to expand the guidelines for his staff and to give you guys some examples of things that you can quote to companies coming to you.

So I've tried to think of it structurally and it seemed to me the first thing was whether there was controversial biology underpinning the observation. So today we didn't hear controversy; there are different companies, they are competing on the same turf, and they have the same conclusion about the underpinning

biology.

But it occurred to me that a couple of good examples where the biology is less clearly defined goes back in time a little bit, but, for example, when we were in the era where MDR drove everything. And you might think in terms of the fact that multi-drug resistance, in its expression, could be responsible for exporting a drug from a cancer cell, at the same time, was being induced by prior treatment.

So the understanding of how to interpret that quantum of information, MDR, plus or minus, whatever that meant, would be controversial. An analogous situation, much more relevant today, is trying to understand the implications of p53 mutation, which clearly has both prognostic and predictive significance, depending on how you frame it, what particular tumor type you're talking about and so on.

The second thing that we have covered today is controversy in the analytic technique. And I think if we go back to the breast cancer literature, if we think about the facts between tumor pathologists arguing about the benefits of fish versus

immunohistochemistry, and different people saying, "Well, my immunohistochemistry lab does a better job than his fish lab" and so on, that's one of the reasons for having the tissue available so you can interrogate it.

But where there is extant controversy in the literature, I would say that is a good red flag for the agency to be much more cautious in terms of interpreting submissions to them, and there, again, I think need for federal standards of what you guys think would be acceptable, because you can say to applicants, "Well, whether we're right or wrong, this is the hoop you have to jump through."

The third thing, I think, again, remembering I'm answering the question of what would be unacceptable, small population size, and, particularly, remembering that in small population size, you can get false negatives just as easily as false positives in big sample sizes. And if you do multiple interrogations or multiple analyses, particularly in small population size, there's no rule that says you only get one false result in a trial. You can get a

couple in a row just at random chance.

I think a thing that hasn't been talked about very much today is the concept of too many variables, the extent of population heterogeneity. Increasingly, we're going to run from one targeted therapeutic to multiple, and we saw an example today of bevacizumab having an unexpected some sort of interaction with an antibody. Who knows what that was. But right there, there as a potential red flag of having two novel agents with multiple mechanisms of action. And so I would see that as a red flag. Using the KISS principle, at least until we understand what we're doing, I think it is something the agency should be looking at. Keep it at one marker, keep it at one agent, looking for interactions where you might hope to see them.

Something that wasn't mentioned today which I thought was actually kind of interesting -- it was put up there, but the applicants -- the sponsors didn't touch on it -- was the implication of adjuvant therapy. And it goes to the point that Rich Simon was making about the stratified randomization doesn't solve all

problems.

So there was stratification. There was even balance generally in terms of the number of patients in these studies who had had adjuvant treatment. And, yet, if you take a look at the results in those populations, in two of the studies, CRYSTAL and 20020408, there were actually differences identified among the adjuvant patients. Now, who knows whether they were significant, because the numbers were too small. But once again, potentially saying let's keep the populations very clean will be important.

The other thing will be heterogeneity in laboratories. Oftentimes, people will say we did a p53, without defining that it was done in one reference lab. It doesn't have to be at the FDA, but at least having one laboratory. So the opposite of that is if you have multiple labs or lack of definition, that would make this unacceptable to me.

Uncertain ascertainment and selection bias. So, again, the issue of what is the downstream impact of having adjuvant therapy. You can turn tumors into expressing p53 mutation with certain types of drugs.

If you have small tumors, this was mentioned, in breast cancer, they can be hard to biopsy and right there you've got a bias coming in. If you have high cellular turnover, you'll get potential necrosis that will confound the biopsy needle, depending on the technology you're using. So a needle biopsy versus full sample biopsy, and even things like prior treatment that isn't in the adjuvant setting. So the patient with prostate cancer who has had castration versus one who hasn't may well have an impact on some of the biological markers we've been talking about.

Then, finally, I think one of the red flags to advise your staff about will be the difference between biological and clinical relevance and significance versus statistical significance. And so you'll often get a P value to the 0.0003, but it may still be talking about a biological effect that's not important. And so I think it comes back to what I said before; using common sense as a goal may be very important.

DR. DUTCHER: Dr. Mortimer?

DR. MORTIMER: This is a different

ascertainment, I suppose, but I would, again, argue that when doing a trial in advanced disease, that a goal should be set to do the biomarker on both the primary and the recurrence to keep from the same problems that happened with Herceptin, which happened with many agents that alter the initial nature and biomarker, the primary, as time goes on.

No matter how much biology we understand, when you treat people, things happen.

DR. DUTCHER: Dr. Simon?

DR. SIMON: Maybe it goes without saying, but to me, one of the most important things is that the analysis be focused on a single scientifically supported biomarker, as others have sort of emphasized.

I don't want to beat that dead horse, but one thing the FDA could potentially do, and maybe it would be done anyway, would be that if a sponsor plans such an analysis, that they would have to clear and have an analysis plan and let the FDA know about that before they actually do the assay. That way, the FDA would know, presumably, how many such markers they potentially looked at.

DR. O'NEILL: Yes, I think that's probably a good idea, but I think it's impractical, because I think what's going on is folks are searching, searching, searching outside the trial.

One way of thinking about this is, even in your work, is if you said, "I haven't decided what my marker is yet, but I want to reserve a part of my uncertainty in my overall hypothesis for that marker," so essentially split your overall alpha into the overall treatment group and then the subgroup, yet to be defined, because all that searching is going on.

That, I think, mimics what's going on, at least today. Now, whether we would move to a different space where someone would commit to us much earlier in the spirit of what you're saying, I think that might be worth thinking about. I don't know how we would be able to pull that off, practically speaking.

DR. DUTCHER: Dr. Funkhouser?

DR. FUNKHOUSER: It seems like studies that have too small a sample set in the subgroup in the particular arm of interest don't lend themselves to demonstrating a difference, even though one may exist.

And so there must be a statistical term for that and I'm just wondering, from the statisticians in the group, how you describe that lower limit which would make the study unsuitable.

DR. HARRINGTON: The term is power and if it were a prospective trial, typically, one would like to see 80 percent or 90 percent power in advance. But there, in a prospective trial, you can set enrollment goals.

In a trial that's been done, the power is not mutable. So I think what I would urge is that people have an understanding, if it's too small, that they probably are going to reach the sort of foregone conclusion of no effect because of lack of power or because of no effect, and you can't distinguish between the two.

If you had a very large collection of samples on multiple trials, you could say we can look at a subset of these samples, because we would have sufficient power in that subset and still be able to do a cross-validation.

One can do those calculations. They vary

from trial to trial. They vary from effect size to effect size. So there's no single number, but typically, one would like to be 80 percent or 90 percent likely to see an important effect, if it exists.

DR. D'AGOSTINO: Again, in our response to topic one, we didn't emphasize that the study that is being put forth, the retrospective study, does have reasonable power before one sort of proceeds with it, not only for the overall, but also the consistency among the different secondary events and subgroups. I think that's very important in putting these studies out.

DR. DUTCHER: Dr. Wilson?

DR. WILSON: I just wanted to make the point that when you're thinking about power, it's a power to distinguish a certain percent difference. And so just to bring up the fact that you can have a two percent difference with a very high power, if you have enough patients, but that may not be clinically relevant. So I think when you're thinking about power and the ability to rule something out, you also have to -- what

you really want to do is you want know the power to rule out something that is clinically significant.

DR. DUTCHER: Dr. Zhou?

DR. ZHOU: That actually has something to do with the definition of the predictive biomarker. So when you say this biomarker is predictive, I'm not clear what you mean. If you only look at the P value, that doesn't give me any information. This is the P value, 0.0001, but it doesn't mean this marker has predictive value.

So the P value is only related to the known hypothesis to say no power. I think maybe we should have some discussion and talk about what the effect size should be in order to say this marker has predictive value, because that's also related to the power, because you can choose any number you want. You can calculate power. But you have to calculate power related to the alternative hypothesis, which has clinical meanings.

DR. D'AGOSTINO: Isn't that what you were saying, though?

DR. DUTCHER: Other comments? Yes?

DR. ZHOU: I want to raise the issue, which I haven't seen that, is to say what's the impact of the subgroup defined by the biomarker, which it has error associated with that.

So those kind of subgroup analyses are different from the typical subgroup analyses you have done before. So before, let's say, we have race and the race variable is well defined. So it doesn't matter if you do it today or tomorrow, you have the same race variable. But for biomarker subgroups, it's totally different. If you define a biomarker subgroup today, compare it with one year later, you may have different subgroups. So that should have some impact on all the analyses we are doing. I don't know the answer, but I think we probably should consider that issue.

DR. DUTCHER: So we're at a crossroads. We have one more question. Shall we just proceed rather than taking a break? Yes? Okay.

Topic four, Dr. Simon and Dr. Grem. When it is acceptable to limit future enrollment to a biomarker selected subset of an actively accruing clinical trial

based on external information?

What would be the primary analysis population? Would the answer depend on the proportion of unselected patients, i.e., those enrolled prior to the study modification?

Dr. Simon?

DR. SIMON: Well, when was it acceptable? I mean, there's no simple algorithmic answer to that. Usually, it's acceptable when the data monitoring committee decides that, ethically, that's important to do.

For example, in the KRAS, usually, it's a big deal when something like that happens. And if there is external information that bears on patient safety and the issues of whether an ongoing trial should be continued in the way it is, it's the responsibility of the data monitoring committee to weigh that information. And they are the right people to do it because their responsibility is to the patients, not to the sponsors, not to the investigators, but to the patients. And anybody else might have some conflicts in terms of what their responsibilities are. So I

think it's appropriate that those kinds of decisions be essentially at the level of a data monitoring committee.

But anyway, when that kind of thing happens, I guess I can envision two kinds of trials that would be ongoing. One would be a trial in which the biomarker-defined subset was sort of something that was known at the outset of the trial and was actually incorporated in the design of the trial; maybe because people originally didn't think that the test negative patients would benefit from this drug. But that kind of information is never for sure, and so never has complete confidence in it, so it was decided to go ahead and include the test negative patients.

Then if some other trial or trials provide relevant information, it may be -- in that kind of a situation, restricting entry probably is not too disruptive in terms of the analysis of the trial, because presumably, in that situation, the trial should have been designed and had a primary analysis plan that included that predictive biomarker. Either it was targeting adequate numbers of test positive and test

negative patients through separate analysis and handling multiplicity and that sort of thing. So I think that's not so disruptive and one doesn't have to rethink, well, what should the analysis plan be now.

I think the more difficult situation is like the situation maybe with KRAS, where the information comes up and it's information that was not available at the start of the trial, and so the trial was not designed with that as sort of the predictive biomarker.

I'm afraid I don't have any sort of great rules of how you deal with it. I think some of these things, there are no rules. They have to be dealt with on an individual basis using sort of the best judgment available.

I think if something as important as sort of external information leading you to discontinue sort of a biomarker negative subset came about, it would be sort of somewhat ridiculous to sort of ignore that in the analysis of the trial. So I would think that probably for the trial, you're going to have to look at the effects overall and also for the positives and for the test negatives.

DR. DUTCHER: Thank you. Dr. Grem?

DR. GREM: So I thought that the situations where it makes the most sense to stop a trial and modify it to exclude patient population that previously would have been eligible would be in those settings where you have information that a biomarker would predict either a patient is at extremely high risk to have harm. So maybe, for whatever reason, it's found that they're not able to metabolize or deactivate the agent that they're being given. So if those patients are treated, they have a very risk of toxicity, or in the situation where if you have that biomarker, then you have no chance of benefit. Those are pretty clear cut.

I think the things that would be much more difficult would be where you're trying to say, "Well, we think that patients may be more likely to benefit." I think that would be very difficult because in all of these things, it's always easier if you can say "You have no chance of benefit" versus "you may benefit," and when you may, you don't know how much that actual benefit is. There's always going to be a lot of other

reasons why a patient may or may not respond to a combination regimen or even to a monotherapy.

So I think it's sort of what Rich was saying. It's kind of a safety issue that if you have no chance of benefit, then the risk becomes unacceptable, or if the risk is so great, then you're unlikely to benefit because you're not going to be able to tolerate the drug and you don't want something bad or a fatal reaction to happen to the patient.

But I think all the other areas, like, "Well, we think that you might be more likely to benefit, so let's not randomize those patients," I don't know that that would really be good.

DR. PAZDUR: Let me just give you some idea of what some of the concepts were thinking about this question. And again, we agree that the data monitoring committee has the primary responsibility of this, but we wanted to get down to some level of granularity here rather than just saying it's the data committee's responsibility.

That is, for example, what was the endpoint that one used to make this decision from the external

information? Was it from only one trial? Was it from six trials? That's going to have a major difference.

What was the effect of the endpoint? Was it a 50 percent doubling in progression-free survival? Was it a six-week improvement? Was it a five percent difference in response rate?

Then, also, if you have this external information that was done in a different disease setting, for example, a very refractory disease population, what implications might that have for another disease setting, such as in an adjuvant setting; should those trials be curtailed and changed, or a first line setting?

So there's a lot of complexities here that one could take a look at in making these decisions and we thought it would be an interesting kind of conversation to have, because as you can see, there's a high degree of subjectivity that could come into play here. There are many factors here that could be looked at and different people taking a look at this external information.

But here, again, some of the issues: effect

size, endpoint used, this constant issue that we've brought up, and I'm happy that the committee also caught onto it, the concept of replication, how many trials this was done; what implications does this have in other diseases potentially or other disease settings in the same disease.

DR. GREM: So at least after going to ASCO and hearing the -- the trials that really struck me as far as the KRAS were the third line studies of monotherapy versus best supportive care because, basically, those studies, I think, pretty convincingly and almost were identical in the sense that patients who had mutant KRAS had no benefit in terms of progression-free survival compared to best supportive care. I mean, they just completely overlapped, whereas there was a pretty big separation for the patients with wild-type KRAS, saying that they would benefit.

When we're talking about a mutation in the gene, I think that if the patient's original tumor had a mutant KRAS, it's unlikely that they're going to regain a normal KRAS. So I think that if they have a mutant KRAS in the primary tumor tissue, they're always

going to be KRAS mutant and we can argue about what percentage or how many cells. That I don't think we have any information to base that on. But I think that was pretty striking data.

So in light of that, I thought that for the CALGB study that's being done through the clinical trials support unit, it made sense to go ahead and stop accrual, modify that, so that only KRAS wild type would be randomized to receive cetuximab.

But the issue about, well, what do you do with the rest of the trial, then, I think that -- and I don't have any control over this, but I would still think that when the trial -- so they increased the sample size so they could now look at the effect of the benefit of cetuximab in patients who are wild-type KRAS, with or without.

But I think that when they finally come to analyze the data, they should look at the original hypothesis, and that was to look at the overall effect in all patients and then to do the secondary analysis in this expanded trial, where they're restricting the analysis.

The things that I don't know about are like for the adjuvant study, then. That study was modified so that if you're wild-type KRAS, then you would be eligible to be randomized to full FOLFOX alone or FOLFOX with cetuximab. And if you're KRAS on mutant, then you're just sort of kicked off the study kind of thing. And I wonder if maybe those patients should have been registered to FOLFOX and followed, because they might be able to provide some balanced information. But I don't think that was done and I don't know why we can't undo that because those decisions were made between discussions with CTEP and, presumably, the sponsors and the investigators who were involved, the primary investigators for the study.

But the questions, in my mind, though, are about, well, is the wisdom throwing the patients off the study and not having information or just they don't get cetuximab, but they can still participate in the study. And I'd appreciate some comments from the statisticians. That's just from a clinician standpoint.

DR. DUTCHER: Dr. Link?

DR. LINK: I think there's a practicality issue, too. I think that one of the things we saw from one of the sponsor presentations was how rapidly this information was disseminated.

Patients aren't stupid. So they may not read it in peer review, but they read it in the *Wall Street Journal*. And if you have a study that shows that you're not going to benefit patients, why would they participate in the trial?

This is back to Rich Simon's original thing that the clinician has to go to a patient to convince them to participate in a trial where they're very unlikely to benefit, and it's difficult enough for the clinician thinking that their data monitoring committee sort of said to go ahead and continue randomizing patients, but the patients may not want to get randomized. So I think there's a practical consideration. The trial may die of its own accord, no matter what the data monitoring committee wants to do.

DR. DUTCHER: Dr. Wilson?

DR. WILSON: So I want to get back to what Rick said, and that is that do you have the same

threshold for applying these kinds of results that were found in the relapse setting to the adjuvant setting.

I think you have to be very, very cautious, because I think that whereas the effect of there being less benefit almost certainly would track into the adjuvant setting, as well, it may not be zero. And if you're actually trying to cure people, the amount of benefit you're willing to accept toxicity for goes higher. So I guess I would have been very cautious to have thrown people off on the adjuvant trial who had KRAS mutations.

We also know, with other drugs, that as you use them in more and more refractory patients, they work less and less well. So again, I have no doubt that they wouldn't benefit much, but it may not be zero. And I think that if there's a difference between cure and not, you have to be very cautious.

DR. DUTCHER: Yes?

DR. D'AGOSTINO: Again, in terms of some of the implications of this question, we've been involved in studies where the results have come out, and I'm thinking more in the cardiovascular, and it changes the

study that you're dealing with not only in terms of a particular group, but also in terms of the endpoint starts changing, and you start recruiting new centers.

So what happens if your original endpoint was something like -- again, in the experience I'm talking about -- an MI, but now you enlarge it to MI and stroke and you end up with new centers and you add on new centers? In the previous centers you were dealing with, you weren't looking for particular endpoints and you weren't following up.

So there are some tremendous potential implications in terms of what has your analysis said that you can deal with at the end of the study and how do you actually make adjustments for that. I don't have any answers, but we've lived through things of this nature, and it's very uncomforting and disquieting to try to figure out what is it that you're going to actually be analyzing and on whom you're going to analyze it.

I think the more we sort of think about this type of question in this setting -- and I'm not so sure these other fields actually bring you much

enlightenment, because they recognize the problem, but don't necessarily have good answers.

DR. ZHOU: I'd also just follow what Ralph said. I wonder whether we can consider reliability of the endpoints. Take the example of the progression disease free versus survival. Survival is harder to measure, but for the disease free, how actually can you really know exactly the time where you have disease free? The measurement is less reliable than the total survival time.

DR. D'AGOSTINO: Well, the things that I'm talking about, you're shifting from an MI and now you start looking -- well, you're shifting -- one of them was cardiovascular deaths and then we shifted to MIs. We weren't collecting all the data to actually diagnose MIs. It was there, but it would have to be done retrospectively, trying to put it all together. The endpoint became a different endpoint and, certainly, overall survival, now shifting it to progression-free survival is going to have serious implications.

DR. DUTCHER: Dr. Netto?

DR. NETTO: I want to pick up on what

Dr. Wilson was saying in terms of excluding, terminating accrual based on data from other settings, like adjuvant.

I think that shouldn't be done for another reason, too. Don't forget, like in this example that we're studying, you have all the other pathways, the mTOR pathway, other markers that, in two different settings, could be totally different. And that probably is playing a role, given the fact that it's only 20 percent of even the non-mutated are responding, so what's happening to those 80 percent, which also brings the issue of -- I know you want to focus on one marker at a time, but probably, in this setting, the other markers should not be ignored.

DR. DUTCHER: Dr. Simon?

DR. SIMON: Well, I think, actually, Dr. Pazdur is right that it's a very complicated issue. It's actually more general, too, because the issue could be generalized to not just stopping enrollment of a subset, but changing an analysis plan without stopping enrollment.

In other words, you may have some information

from some other trial which may not want to make you necessarily stop enrollment, but the trial you were originally doing may not be -- that analysis plan may no longer be the most relevant analysis plan. And so then you get the issues of is it okay to change an analysis plan. And changing that analysis plan may involve increasing target accrual rates for subsets and things like that.

So it gets complicated. I think my only -- and again, on a lot of these things, I don't think it -- I know that guidelines are useful, but I think, in many of these complicated situations, guidelines only carry you so far, because there's too many different situations.

But I think there probably needs to be somewhat more recognition of the actual relevance of potentially changing analysis plans prior to analysis of the data, when the data is still blinded, but at a time after the study has started, because of the complexity of developing biomarkers.

DR. DUTCHER: Dr. O'Neill?

DR. O'NEILL: I wanted to follow-up on this

and sort of revisit why I showed that slide this morning by Professor Moyé, which were the three examples in the cardiovascular area, where they were fooled.

And Ralph is correct, because Ralph has sat on cardiorenal advisory committees and see this in action, where a lot of smart people changed the endpoint midstream and, at the end of the study, it lost. And if they hadn't changed it, it would have been okay. They would have won.

So there are many examples of this and I guess my concern is, even at best, that's just one study anyway. So you're fooling around with one study and you're making some midcourse changes, where you really don't even have a good analysis plan, because the question originally came about as how do you count the individuals that you've already accrued in that are now marker negative and you're not going to accrue anybody else in there. And everyone was saying, "Well, you actually have to keep them in the analysis." Sure, that makes sense, but there are other issues that are going on here which have opened the door for possibly

changing other aspects of the trial.

There's another interesting wrinkle to this and we're seeing this in multinational, multiregional studies. So think of the modern clinical trial being done in three regions, the United States, North America, South America and eastern or western Europe. And you're sitting on a data monitoring committee and you're starting to see patterns and where there is possibly no effect going on in one area.

The question is do you -- the analogy here is there's no effect. Do you stop accruing in that area, saying it's unethical to continue, or do you essentially say something is going on and I don't know whether this region is going to be able to share in the overall effect that I'm seeing maybe outside the United States.

We have a version of this in results in other areas, where the results either looked better or worse in the United States versus outside the United States. So this is a version of subgroup analysis and not knowing what's true and what's not true. And you certainly complicate it if you change the endpoints,

but I'm not even at the point of changing the endpoint. Mid-trial changes have a history of fooling a lot of people in many areas, and that's just sort of the general message.

DR. D'AGOSTINO: I've seen studies where you actually throw out the centers that aren't producing any events, and what do you do with them at the end of the study? You have the data. What do you do with it? Do you throw it in or you ignore it?

DR. DUTCHER: Dr. Simon?

DR. SIMON: I would agree that these issues are complicated and they have to be dealt with carefully and not willy-nilly. But I think we do have to bear in mind that the kinds of predictive biomarkers we're talking about are not retrospective data dredging kind of biomarkers. And I think, to me, oncology is actually leading the march to personalized medicine and predictive medicine.

I think we are already here in oncology and the question is to try to make sure that studies are done well and the regulatory environment is conducive and encouraging to those studies and for sponsors to

develop these biomarkers. And I think what we will find, actually, is that these other fields of medicine will wind up following oncology. And so I think they have, actually, a lot to learn from us, more so than the reverse.

DR. DUTCHER: Dr. Pazdur?

DR. PAZDUR: I think one of the areas that we were very interested in, in talking about this external information, is how robust this information is, because here, again, we do have this tendency of a slide toward the least common denominator here. Here, again, it has to do with effect size of the external data. It has to do with the reproducibility and what endpoint is being looked at.

For example, we have a lot of experience now with interim analyses of PFS data and there could be a high degree of fragility with this endpoint based on interim analyses and looking at what the expert review committee has to say about it and measures it versus the investigator when you're talking a look at an interim analysis.

So one of the major issues here and one of

the reasons why we asked this question, we really wanted to have a discussion, and I think we have had that, looking at this should be an effect that people feel comfortable with here. And as Dr. Simon pointed out, there needs to be confidence that we have this effect before we stop trials, et cetera.

DR. DUTCHER: Dr. Netto?

DR. NETTO: So the question is, again, just to pose it, so how many trials you need.

Would the two trials, prospective/retrospective trial, well conducted, be enough to point in one direction to stop accrual?

DR. PAZDUR: Can I give you the FDA answer?

DR. NETTO: No. I want your answer.

Dr. Simon?

DR. DUTCHER: What did you say?

DR. PAZDUR: I was going to give him the FDA answer. It's a review issue. And it really depends on the number of trials that you have, the magnitude of effect, the persuasiveness of that effect statistically, consistency within trials.

So to say I'm going to give you a number is

reminiscent of sponsors that come to us for accelerated approval and ask, "What is the lowest response rate that you will take?"

DR. NETTO: That's why it has to be two well conducted studies.

DR. DUTCHER: Dr. Raghavan? Dr. Wilson?

DR. WILSON: So we've been talking about stopping trials, and I don't know if this is the place to ask this. But now that we are circulating around this idea that having wild-type RAS is what counts and knowing that immunohistochemistry for EGFR is a very slippery slope, are the companies planning on looking at wild-type RAS rather than EGFR in colon cancer and looking to see whether or not there is a benefit in EGFR negative by IHC than wild-type RAS positive?

DR. REESE: Davis Reese from Amgen.

In our ongoing Phase III trials that I presented to you earlier, the 181 and 203 trials, those trials do not actually require EGFR as an eligibility criterion. We will be doing EGFR staining on all of the specimens, hopefully, nearly 2,400, and we'll be performing a variety of analyses to correlate outcome

with EGFR expression. We're working with the agency on those analyses.

DR. NETTO: No fish, for amplification?

DR. REESE: Gene amplification by fish is extremely rare in colorectal cancer as opposed to lung cancer or other diseases.

DR. DUTCHER: Dr. Funkhouser?

DR. FUNKHOUSER: Two comments. It seems that if you have a big trial and you stop accrual because you're convinced that there is no potential benefit for patients with some particular genotypic variable, then you send a message to the academic and the commercial communities that no further research in this area is necessary, I think it's unlikely that you're going to get large trials that replicate what you have ongoing. So the cautionary tale there is that it seems that you need to be statistically right, as well as emotionally confident.

The second point is that just because RAS is wild type doesn't mean that other proteins in the same signaling pathway, think BRAF, aren't mutated, and those may be some of the non-responders. Remember,

86 percent of your patients are non-responders that are wild-type RAS. So some subset of those may be other mutant signaling proteins within the same signaling pathway, MAP kinase and BRAF. And we've seen this in thyroid carcinoma, where RAS and RAF are separate complementary subsets, either of which can be mutated.

DR. DUTCHER: Dr. Youssoufian, did you want to comment?

DR. YOUSSOUFIAN: Thank you. So part of our post-marketing commitment for the initial approval of Erbitux was to perform an EGFR negative study in refractory colon cancer, and we've actually completed that study and presented initial results at last year's ASCO. There appear to be a handful of responders. It's a relatively small study, about 80 patients, so it's hard to be somewhat more quantitative about it.

But for all intents and purposes, at this point, we're -- and not just we, but I think the general community is regarding EGFR negative and EGFR positive, at least by immunohistochemistry, as essentially the same group.

So to do a KRAS study in those two different

groups will have to have another biological hypothesis.

DR. DUTCHER: Dr. Richardson?

DR. RICHARDSON: It seems to me that a lot of this discussion is predicated on the assumption that the biomarker and the biology of the underlying disease are independent, and we certainly saw some data earlier suggesting that, for example, the folks with the wild-type tumors had the same type of survival as -- given best supportive care, had similar survivals to the mutant KRAS group. At the same time, I guess we also saw some data that suggested that in the Pmab studies, the patients with the wild-type tumors actually had better survivals than the mutant KRAS population.

I'm just wondering whether we can account for some things that may be even more subtle, though, whether this enters into this.

What if you have a biomarker that is more associated with, say, oligometastatic disease? So that at the end of the trial, where the patients have all progressed at a certain point, suddenly, in this group of oligometastatic patients, the surgeons say, "Well,

you know what? We can take out those nodules" or "we've now got RFA for these hepatic mets. We can cook them or we can chill them with cryoablation," and suddenly, in that group, your overall survival figures are going to change subtly, but maybe enough to shift the curves.

I think we've got to figure out some way of dealing with these kinds of changes in medicine, as well. This is something that is happening around the country. Surgeons are becoming more aggressive than they were five years ago, ten years ago. Interventional radiologists certainly have the ability to deal with some of these lesions in a way that is much more effective in terms of de-balking these patients than we ever were able to do previously.

Radiotherapists are now saying, "Well, you know what? We can treat those lung mets" or "we can treat these other lesions elsewhere in the body using the CyberKnife," for example. All of these things are going to impact these overall survival numbers just by de-balking some of these people, and it will be enough to shift these curves.