For reasons already discussed, such well choreographed development of a new IVD test along with a new drug is a rarity so far.

When biomarker information emerges as part of exploratory analyses, the most informative way to move forward is with a prospectively designed randomized controlled trial. This approach assures several study strengths that become questionable when retrospective mining of previously banked trial specimens is used instead.

First, with the prospective trial, the relevant biomarker is pre-specified and there is no worry about inflating test performance through multiple comparisons used to identify the biomarker.

During specimen accrual, specimen collection, preservation and storage can be optimized for assay of the biomarker. With prospective design, specimen accountability may be substantially increased and non-random loss of specimens, with its attendant risk of bias, can be actively minimized.

In executing the new trial, there is the opportunity to stratify on the biomarker status before

randomizing to treatment groups. In designing the new trial, one has the opportunity to manage the allocation of Type I statistical error or to maximize trial efficiency through adaptive design. These are the strengths that should be weighed against the convenience and insights available from retrospective analysis of clinical trials. The next presentation from Dr. O'Neill will explore these tradeoffs in greater detail.

So to summarize, companion diagnostics are at the heart of personalized medicine and carry the same risk profile as the drug. Predictive claims for companion diagnostics rely on understanding the effect of the drug in both biomarker positive and biomarker negative patients.

Late emergence of critical biomarker information causing reevaluation of a well studied drug in light of a new biomarker may become common. Because of this, analytical validation of the new IVD test may occur late in the development process, but it is essential to complete it before testing clinical trial specimens.

Randomized controlled trials have substantial advantages in evaluating the significance of late emerging biomarkers. The extent to which revision of the drug's use and clinical validation of the IVD test can be based on retrospective analyses of retained specimens requires scrutiny, and this will be the topic of further presentation to follow.

Thank you.

DR. DUTCHER: Thank you.

We're going to move right along to Dr. O'Neill's presentation.

DR. O'NEILL: Good morning. I'm Bob O'Neill. I'm the director of the Office of Biostatistics, and I've been asked to talk about this problem because it's an issue that just does not arise in oncology in this KRAS situation, but it's also an issue that we're seeing across the board in many drug disease areas. And this is a great opportunity to discuss and get input from the committee about what should be the general principles as we go forward in this area.

So this is an outline of what I'm going to talk about. I will repeat some of the concepts that

Dr. Becker talked about, which essentially are definitions of prognostic and predictive, and I'll go over some general principles for the design and analysis of clinical trials for subgroup differences, particularly the issue of the control of the false positive and false negative conclusions and the issue of subgroups defined by pretreatment baseline characteristics rather than some characteristics that are ascertained while you're post-randomization and may be influenced by therapy. But that's not what the issue is here.

I will take a shot at what we're now calling prospective/retrospective genomic clinical trials and I'll spend some time on the principle of replication, the idea, even under sort of the substantial evidence principle, of how much evidence do you need to demonstrate a subgroup finding, not only an overall effect, but a subgroup finding.

And we'll talk, as illustration, about some of the KRAS studies. We have six of them that Dr. Giusti had described for you a little earlier and I'll return to some of those. And then I'll go through

some prospective study designs that are available and have been published and are available, which would not essentially be called retrospective designs.

So we made the following comments to the sponsors with regard to what we are dealing with right now. And the optimal approach is to conduct a prospective adequate and well controlled trial that's designed in advance to assess the subgroups based upon KRAS testing by validated assay, or another pragmatic approach, which would be a retrospective analysis of the studies that have already been completed or are ongoing but under the following conditions: that these studies be adequate and well controlled; that they be a large enough sample size; that the factors that were not stratified on as randomization factors, in particular, KRAS status, would balance out; and, KRAS biomarker status was ascertained in virtually the entire population, let's say, greater than 90 percent of the population -- that's not the situation that we're dealing with with the majority of the other studies here -- that the assay be acceptable and have acceptable analytical performance and we have an

acceptable analysis plan.

So let me start, first, so we're all on the same page, with regard to the definitions.

We're using the word "prognostic" to indicate a marker for which the magnitude of the event or an outcome is related. So it's independent of treatment. So, essentially, the way we've used this in the past is to enrich a population for an event because we'll see more events. Then there is the term "predictive," which is what we're primarily focusing on today, which depends upon treatment status and is relative to a control group. So this is defined as a marker for which the magnitude of the treatment effect is related.

So here is a scenario. A, you have marker G- and G+, are the positive and negative status of the groups that are ascertained by the assay. And what you have here is a control group where there's a 33 percent response rate in the G- and the G+ groups. So the outcome itself is not differential in the non-treated group. But on the drug, there is a 15 percent response difference on drug A in the G+ group and no response in the G- group.

That is called a predictive treatment effect and it's seen only in the G+ group and, in fact, statistically, we call that a qualitative interaction. That is to say there's no effect in one group and there is an effect in the other group.

We'll move on to scenario B, where what you have is a differential outcome in the control group in G- and the G+ group. So you have 39 percent and 48 percent. This could be a response rate or whatever, cure rate.

What you have here is no effect in either group, in the G- or the G+, but you do have a difference in the magnitude, and you could change that and essentially have a difference in the drug group, also, but no treatment effect, a ten percent difference in the G- group, a ten percent difference in the G+ group. So there's no difference in treatment response.

So that's essentially the way we're defining it as a prognostic marker. So there is an effect in G+ and G-. They are consistent. There's either no treatment effect or the same treatment effect, but it's differential.

Then, finally, you have the situation of scenario C, which is where the control group has a different background rate than in the G+ and the G- groups. There's 39 percent versus 48 percent. And there also is a differential treatment effect in the G+ and the G- groups.

So both groups share a treatment effect, but to a different degree. And so what we're calling that is prognostic or predictive, where the effect is larger in the G+ group than in the G-, but they both share an effect. And this is called the quantitative interaction, relative to the earlier one, which is a qualitative interaction. And we're interested in both of those situations and the issue is what do we need to rule out there is no effect in the minus group.

I'm not going to spend much on this, but as a general issue, if we were not talking about KRAS -- and Dr. Becker did describe what are the minimum performance characteristics for a marker. And you might consider that in terms of its sensitivity and specificity and other characteristics, because the ability to classify a patient repetitively the same way

is an important issue.

What are the consequences? Because if we were talking about other markers, there might be other implications. But in general, if you have five or six or seven studies and you're trying to see whether they are all telling you the same thing, they could differ primarily because the classifier doesn't have good sensitivity and specificity. So a combination of poor sensitivity and specificity and a different mix of that classifier prevalence in those studies could contribute to why you might have heterogeneous results across studies. So that's why it's important to take a step back and look at that when you evaluate multiple studies against each other.

So when we said we would like a good analysis plan, this is what we were thinking about: the role of randomization to assure unbiased and fair comparisons; the role of marker status classification, the impact of convenience samples on biased estimates.

So what we're talking here, if we were in the situation where every one of these studies had gone back and looked at 100 percent ascertainment of marker

status, that's one issue -- we're not in that game. We're into a situation where anywhere between 20 to 90 percent, 95 percent, of the ascertainment of the samples, and this is an issue that I'll return to in a moment.

Then we have the statistical control of the false positive conclusions. So we worry about how many hypotheses are tested, whether you tested one and then you went back and tested another after you found out one didn't work or whatever. So this is very much about primary analyses, which failed, how many ways can you win, how many outcomes are there.

We're talking about outcomes that are overall survival, progression-free survival, and response rates. So depending upon which was a primary endpoint and which is now being looked at in terms of its impact and relationship between KRAS status -- it's just the general issue of multiplicity and data to generate the hypothesis versus data to confirm the hypothesis.

So our general strategy in a clinical trial community, in general, has the following strategy. You look for an overall treatment effect on the primary

outcome in the intent to treat population. That's everyone who is randomized. And if there's a statistically persuasive result, generally, that means if the Type I error controlled at five percent, the P value is less than 0.025 one-sided, then you can do other things with the trial. And you can look at subgroups and there are various options. You can look at marker negative group, marker positive group. You can examine for equal treatment effect, interactions. And that's generally the approach.

If there's no statistical significance on a primary endpoint, that's primary hypothesis, everything further is exploratory, not useful, but exploratory and generally is useful for what you do next.

If along the path, you change the number, the set or the sequence of hypotheses after the start of the study, that's generally not considered an acceptable practice, particularly after you've observed the data. But there's a lot of interest in adaptive designs these days, which essentially pre-specify the kinds of adaptations and the kinds of changes that you will make, and those are acceptable.

We've introduced this term, or others have introduced the term, "prospective/retrospective" study. What is it and what are the concerns about it?

Well, you can consider part of the definition being the classification factor is not known at the time of study initiation and the study is, at first, not analyzed with that factor as part of the hypothesis. So this is the retrospective aspect of it. The initial hypothesis and the endpoints for the study are not changed, except if pre-specified as part of a planned adaptive study design. But that's not what we're talking right now.

The control of the false positive conclusion from the studies are appropriately dealt with. The randomization is not stratified on a factor that itself is of interest as one of the hypotheses to be tested. And the factor of interest is ascertained at baseline on all subjects randomized to treatment. And the question is what if that's not the case.

So here's a working definition for this prospective/retrospective design.

In a completed or post-interim analysis

trial, where the genomic samples are collected prior to treatment initiation, whether or not full ascertainment on all subjects, the genomic hypothesis is prospectively specified prior to diagnostic assay testing. However, the clinical outcome data, without genomic information, has already been collected, un-blinded and analyzed. And the genomic data analysis might be arguably prospectively performed, which is, in essence, a retrospective analysis.

So that may be a bit much for you to chew on, but that's what we're talking about, and that's not bedtime reading. I'm just sort of going through it.

So the big issue here is a convenience sample and we're going to return to it. So that's essentially when only those who are around give you the samples or only those who are available, only those who have tissue samples are available.

Another part of a good analysis plan is controlling the chance of erroneously concluding that there is a real treatment effect, when, in fact, it is not, or the chance of concluding there is no treatment effect, when, in fact, one actually exists. These are

two critical issues and we're talking now about subgroups and three endpoints, primarily progression-free survival and overall survival.

The other aspect is the importance of randomization and that's so that you're comparing likes with likes. And this is especially an issue with the interplay between convenience samples, where you don't have the full randomization population and where you may have some characteristics of the assay that might not be well known.

So why this is important, let's look at some data. Here is the EPIC trial. This was CMAB. This study was described briefly by a number of the previous speakers.

And I point out that this was a trial of approximately 1,300 subjects, where 300 subjects have KRAS status ascertained. This is about 23 percent. And these were all from the United States. And the primary endpoint in this study was overall survival. And here is the issue.

There was a dramatic difference between the hazard ratio in the ITT population and the convenience

sample, and the overall population hazard ratio was virtually, one, 0.98. But in that subset of 300 subjects, the hazard ratio was 1.25, going in the wrong direction, for overall survival. So there is a difference, and why there's a difference, one could hypothesize many reasons. But the point here is the difference between the convenience sample and the intent to treat population.

Now, if you look at this by wild-type and KRAS status, mutant status, for overall survival, there is no difference between them, 1.29 and 1.28. But if you look at another endpoint, which was not the primary endpoint of the study, progression-free survival, this is where there appears to be a differential difference. And you have a hazard ratio in the wild type of 0.77 and in mutant, you have 0.10.

So if you sort of look at this and you don't know how you got there and the logic of how you got there, this is all -- I'm trying to deconstruct this particular example, saying this is why we are concerned about convenience samples.

So let's look at all the studies that we have

before us, and we have six randomized studies for which marker status is available only on a selected subset of the randomized study population, ranging from 23 to about 92 percent.

We, at this point in time, have no documented evidence that the treatment groups with ascertained marker status are comparable for baseline variables. We haven't really reviewed that data in detail. And this is primarily to illustrate the conceptual points of how would you look at the data to look for consistency.

So what do we need to know for a marker to be predictive? We need, as I said earlier, an unbiased comparison between the test treatment and the control in each of the marker subgroups. Unbiased generally means a randomized subset of subjects in each of the marker categories, not a convenience sample of subjects with the marker status.

So these were the studies that Dr. Giusti had described earlier and, essentially, these are the CMAB trials and they're outlined in orange. Two of them met their primary endpoint and two of them did not.

We're talking about first line, second line, third line therapy. We're talking about different control groups. We're talking about different endpoints. So the issue of likes with likes comes in in terms of trying to say are these all telling me something directionally.

Now, we have the PMAB trials. There's two of them. And the first one, which was described, met its primary endpoint and that is the one with the 92 percent ascertainment of the KRAS status. But there's another trial which is an interesting trial that did not meet its endpoint. In fact, it was significantly inferior, and I'll return to that.

So let's look at all the studies and stack them up relative to each other and are they telling us a consistent story. And when you think about this, you might think of it in terms of line of therapy, control group endpoints, convenience samples, and remember the EPIC study which I just described, which the U.S. results were very different.

Now, let's talk about overall survival. In this next slide, this may be -- bear with me, but I'm

going to read this slide because it's relevant to interpreting this particular graph.

This graph provides a summary of the overall survival for the five studies that have overall survival comparisons for the wild-type and mutant KRAS subgroups. The hazard ratio is what is describing the effect. Points above the line correspond to larger effects for CMAB or PMAB for the mutant KRAS status and below the line, it points to larger effects for CMAB or PMAB for the wild-type subgroup.

So what this is saying is the circled point is the only trial that shows a greater benefit for the wild-type KRAS. All the four other studies do not show any benefit for the KRAS wild type, for the most part. I mean, you can sort of talk about this guy right here. But essentially, the point is that any point that is on this line means that both the wild type and the mutant type share the same effect size. And in this particular instance, they share the same effect size going in the wrong direction. And in this particular instance, certainly, there's benefit in the KRAS only and no benefit in the wild type.

So this is sort of a summary of -- and I think there's going to be an important distinction between overall survival and progression-free survival. And if you look at this with regard to just the PMAB studies, there is no study; neither of the PMAB studies show any benefit on overall survival for the wild type.

Moving on to progression-free survival, the same explanation for this slide. What we have is we have six studies and five of them are beneficial for wild-type KRAS. So on the PFS endpoint, that is where directionally it looks like the action is going on, and then there's one study here where that was actually significantly in the wrong direction for both wild type and for mutant status.

So the point of this is how do you look at it collectively in terms of is it hanging together.

Well, there are study designs that you could do prospectively, not retrospectively, and they've been published, and you could do a two-stage design that reserves some of the Type I error for testing a subgroup yet to be specified. That could be a fixed study design or it could be an adaptive design, where

you could upsize the trial according to the power you needed in the marker subgroups, or you could test the efficacy of a strategy that screens for the classifier.

Some of you may be aware of this, but this is an interesting trial that was done on the drug Abacavir. It's an antiviral drug. And the problem with Abacavir was it had about an eight or nine, a ten percent hypersensitivity reaction that was very severe and the drug was not being used because of that.

Anyway, for a number of reasons, a lot of historical data, a lot of case control data, a lot of observational data, but essentially it was felt that to convince folks that there was a benefit to screening for someone who would be HLA positive and would have a hypersensitivity reaction, this particular trial was conducted.

So they randomized individuals to the drug and the only difference between the arms was in one arm, they screened out one of the HLA subgroups, and, in the other group, they took all comers. They typed everyone and then retrospectively analyzed those two comparisons.

I'm not going to read the text here, but this study confirmed the hypothesis that screening will reduce severe adverse reactions by a half. So essentially, the screening strategy reduced what was a background rate of eight to nine percent to about four to five percent. And some folks might say it almost reduced it to zero, depending upon how you define the phenotype of the hypersensitivity reaction.

One of the other byproducts of this particular design, in one of the arms, you directly were able to calculate the sensitivity and specificity of the screening strategy. So the byproduct in that particular design was to get estimates of sensitivity and specificity.

There are other approaches. One of my colleagues, Sue-Jane Wang, has written about this and this is not the only publication on this, actually, which would incorporate some adaptive designs. And Dr. Richard Simon, who is on our panel today, has written extensively on this and these are two articles that he has written that are on this particular topic. So there are prospective ways of doing this.

So just to remind you that we have not reviewed these studies in detail. We don't have actually access to the patient level KRAS data and I'm only using these primarily as illustrative.

Let's turn to the issue of the scientific principle of replication. Generally, when you have -- you didn't design the trial in the beginning for a subgroup finding, and subgroup findings turn up, there is somewhat of a caution out there. And this is a *New England Journal of Medicine* article on statistics in medicine, last year, on reporting of subgroup analyses in clinical trials. And, essentially, it was written as a tutorial to say here is what you need to be concerned about and looking for.

Generally, the concern has been that subgroup findings are exploratory, at best, or false positives, unless further evidence is available, and this is essentially prior evidence or other studies. And the strategy has generally been to statistically adjust for multiple analyses, test for interactions, which have been done in all these studies by the previous speakers, interactions meaning looking for a

qualitative or a quantitative interaction. Is the effect size different in the different G+ or G- groups?

So the importance of that is there is this caution with subgroup findings and the issue of if I only saw it in one study, should I repeat it and how often do I need to see it.

And here, we have that issue, just not because it's a subgroup finding, but all of these subgroup findings are from convenience samples, because these studies are not 100 percent ascertained going back into the current study. So you have both of those at play here.

It's interesting to see how the cardiovascular community has reacted to these kinds of issues. If you've ever been at any of our cardiorenal advisory committees, this is an article from Professor Moyé, who sat on our cardiovascular committee a few years ago, and there were a number of examples that came in studies where the results were rather dramatic in a subgroup. But after the sponsor was asked to repeat the study, the results turned out to be essentially negative.

This is an interesting read. The main point here is why would these trials, whose findings were reversed after an original subgroup finding stood out -- was primarily because the analysis plan in the initial study changed after seeing the data. It placed new emphasis on a subgroup finding or on a secondary endpoint raised in prominence leading to false discoveries that were not replicated.

Now, I'm not saying that's what's going on here. I'm just saying that the general issue of how much evidence you need in a repetitive replication concept. And this is taken from the Website of heart.org and, actually, this is a Bob Califf write-up about one of the drugs here, which is PRAISE 1.

This was done in 1,100 cardiomyopathy patients, and the unexpected finding was that the drug was beneficial for patients with non-ischemic cardiomyopathy, a 31 percent risk reduction in the primary endpoint, which is mortality. And essentially, PRAISE 2, which was identical in design, but probably four or five times the size, was done, and the favorable survival benefit of amlodipine, as seen in

PRAISE 1, was likely due to chance, they claim, despite the fact that mortality was an unequivocal endpoint, the benefit was seen in a pre-specified subgroup, and the P value for the subgroup was very small.

So I'm just using this to illustrate somebody else's thinking and approach, particularly, how the cardiovascular folks have thought about this. And the critics argue that although prospectively defined, the subgroup was small, 119 patients. That's the ballpark of some of the subgroups that we're talking about here.

So anyway, enough about that. Back to the questions.

We asked the sponsor to address the adequacy of the analysis plans and the analysis of the data. I believe they've given a fairly good effort at that, what you've heard this morning. I've tried to describe some of the available studies to illustrate the points here and sort of lay out what I think you need to think about in terms of some of the limitations.

And finally, sort of the major question, how much evidence is needed to establish or support a predictive marker claim? In many ways, what we have is

a collection of associations. They are not really external to any of the studies. They are sort of collectively -- I mean, the motivation for how this moves very fast, I think was described earlier. And I think the question before us is is it real and is it repeatable, and what's the role of two independent studies, both of which are prospectively designed, not retrospectively analyzed, to test the marker hypothesis, and what's a reasonable metric for the strength of the statistical evidence, particularly when you think of the effect size, the consistency across studies, the sample size and the subgroups, and the appropriateness of the randomization for the convenience samples.

With that, I think I am finished. Thank you.

DR. DUTCHER: Thank you very much. Very interesting food for thought for all of us.

We're going to be taking a break for 15 minutes and then we'll come back and have some time for questions to the presenters. So it's 10:35. So we'll be back here at 10:50.

(Whereupon, a recess was taken at 10:35 a.m.)

DR. DUTCHER: We're going to be opening it to the committee for questions to the presenters, keeping in mind that in five minutes, we're going to have a fire drill. We do not have to leave the room. It is a test. We just will probably not be able to talk.

Yes, sir? This is Dr. Funkhouser? Oh, D'Agostino. Okay. And please state who you're addressing the question to.

DR. D'AGOSTINO: This is a question to the FDA group.

I'm really concerned, as I'm sure everybody is, with how you start selecting these studies. And one of the comments that was made is that we don't want to salvage a negative study. But it seems to me that's exactly what will happen because the negative study will be declared to be negative because we had the wrong genotypes mixed in and so forth, and what you want to do is to pull out that subgroup for salvation.

Do you really have sort of rules in your mind now that these type of studies, in order to start playing this game out, you can only be dealing with studies that were originally -- prospective studies

that were originally positive?

I just don't see how that's going to materialize.

DR. PAZDUR: Let me address part of that and maybe Bob wants to jump in.

When I made those comments, we were talking about people coming in for primary efficacy claims of an application and we have had numerous examples of this, where people fail their primary efficacy trial and then are coming in with one trial based on, "Well, we failed this trial, but how about if we take a look at this subgroup of patients." And I think everybody understands that that is purely an exploratory area and certainly not something that one could consider substantial evidence that would warrant approval of a drug.

Here, on the other hand, we're talking about somewhat of a different scenario here. These are labeling claims. Sometimes they could be made for safety. We're talking about looking at, basically, in some instances, predefined hypotheses. We're talking about replication of findings in multiple trials. And

I think Bob made a concerted effort to discuss the fact that we want replication of findings, not exploratory analyses in one failed trial and let's approve a drug. And that's the point I was trying to make.

DR. D'AGOSTINO: That's great. So of these qualitative type of interactions, it may be that the overall study isn't positive, but when you drill down to the subgroup, appropriately, in replication and so forth. Thank you.

DR. DUTCHER: Dr. Mortimer?

DR. MORTIMER: So I have a question, I guess, to either the FDA or to our sponsors regarding tissue that was used to assess the KRAS status.

So drawing from, I guess, the breast rule both for ER and HER2, there is change in time of these markers. And so what is the proper specimen on which to do KRAS?

Obviously, in the adjuvant setting, that would be easy. But in the metastatic disease setting, is the primary tumor and the metastatic site going to be the same? And does prior treatment -- so, again, going back to the HER2 status, women who have been

previously treated will frequently, 30 percent of the time, change a HER2 status, which none of us actually expected as those studies moved forward.

DR. KEEGAN: Well, I would say that, just speaking from the FDA, that we really haven't been presented with enough information to answer that question. But I thought your question was actually starting somewhere else, which was even what is the tissue that you -- how do you preserve the tissue, how do you handle the tissue.

(Pause)

DR. DUTCHER: I guess that was the test.

Go ahead, Dr. Keegan.

DR. KEEGAN: So I was going to say that that's part of the workup and the characterization was we would expect to see that kind of information, is it important to know whether it's pretreatment or can you get it anytime; is there drift over time.

But, also, there is a substantial issue with how do you preserve and handle the tissue, is a biopsy as good as, a needle biopsy, as a core biopsy. All those things are very important and the problem is it

depends what the test is, it depends what the analyte is. And when you haven't defined that up front, you don't know whether or not the -- even if you have 100 percent acquisition, whether or not those samples can really be evaluated because of handling.

DR. REESE: If I might be recognized by the chair?

DR. DUTCHER: There you are. Okay.

DR. REESE: Dr. David Reese from Amgen. I'd like to add a little color to your question here and I think there may have been a couple questions embedded in there, number one, what type of specimens are used and are appropriate and, two, how reliable is that testing on different types of specimens?

In terms of specimens, they're typically formal and fixed paraffin-embedded tumor specimens.

(Pause)

DR. REESE: In any event, these are routinely available -- I'll speak quickly -- archived specimens that would be standardly prepared, fixed and archived by a pathologist.

Another question that you may be asking is

how is the stability of the testing on samples of different ages. We've actually taken a look at that in a number of our specimens, for example, looking at them in specimens of different times. And if you look at specimens, for example, less than four years old or more than four years old, the percent of patients with KRAS wild-type tumors is identical in a relatively large number of specimens.

DR. MORTIMER: Actually, my question was even more simplistic than that. We know that when you treat women with metastatic breast cancer, that over time, those women that were HER2 negative often end up with HER2 positive cancers, and who knows whether the same might hold true for KRAS, because certainly nobody expected that with HER2.

DR. REESE: Sure. David Reese from Amgen again. That's a very important point. I think what you're referring to is the stability of this mutation as tumors may evolve in their natural history.

KRAS mutations are relatively early event in colorectal cancer, pathogenesis typically occurring at the aberrant crypt or adenoma stage. Based on the best

data to date of various concordant studies in the literature, they suggest that when assessing KRAS status in the primary tumor as opposed to metastases, that status is concordant in a very high percentage of patients.

What we don't know is the specific answer to your question, which is under potential selection of an EGFR inhibitor, does that change. We are, in fact, performing a biomarker study that will obtain tissue prior to and after exposure to panitumumab and KRAS, as well as other biomarkers will be assessed to determine if that evolves under drug exposure.

DR. DUTCHER: Dr. Becker?

DR. BECKER: So to speak in a general sense, though, I think that your concern with respect to the issues associated with marker stability across the universe of markers that could be faced in this kind of situation is a well taken point with respect to marker stability, especially for immunohistic chemistry, but as you look at other kinds of markers, for example, RNAs, that may be more labile in the pre-fixation stage, where, as many pathologists might recognize, the

idea of a standardized approach to processing the issue immediately after removal might be very difficult to achieve, if it can be. These are variants which could be important in situations that we will encounter in the future.

DR. WILSON: I just wanted to point out that even if the mutation wasn't stable, for this type of study, it would not adversely impact the outcome given the data already shown. For example, it would reduce the effect in the wild-type group, and we've already seen an effect there. And it wouldn't adversely -- and it wouldn't impact the mutated group.

So I don't think that that really, given the data, would have any type of adverse effect on analyzing this data.

DR. DUTCHER: Dr. Harrington?

DR. HARRINGTON: I had a question about the samples, but my question is, is this still the drill? That's my question.

Thank you. I think both FDA and the sponsors have --

(Pause)

DR. HARRINGTON: All right. Let's try again.

So I started to say that both FDA and the sponsors have laid out, I think, a lot of the sand traps here on the way to the green and trying to sort this out. I think one of the things that most intrigues me and concerns me is the extent to which the samples are not fully ascertained moves the analyses away from the analysis of a purely randomized trial and closer to an observational study.

So what I'd like to learn from the sponsors is more about the barriers to getting full, first, acquisition and storage of tissue material in potentially pivotal trials and then being able to archive and use those so that, at least prospectively, that issue may be less severe than it was in some of the trials that we saw this morning.

DR. DUTCHER: Dr. Youssoufian?

DR. HARRINGTON: And how you might solve some of those issues, as well.

DR. YOUSSOUFIAN: Sure. Hagop Youssoufian from ImClone Systems.

So those are indeed the most pragmatic issues

that we also deal with in terms of enriching for the types of analyses that we're all interested in.

I can tell you that on all four trials that I described, the issues for the sample retrieval were very, very similar. So lack of informed consent was probably the number one reason.

Geographically, in one of the protocols, in the EPIC protocol, the second line study, per protocol, it was specified that the samples will only come from U.S. patients. Now, that's something that we would certainly rethink in future trials, to obtain samples from more representative populations.

With regard to storage issues, those were stored in a secure area in a central tumor bank maintained by the NCIC, and I think it was mentioned before that for DNA tests, perhaps the DNA as an analyte --

(Pause)

DR. YOUSSOUFIAN: Okay. I'll squeeze in the last set of comments.

That DNA is intrinsically more stable than protein, but, nonetheless, those are still very

important problems to be cognizant of.

One bit of information that I'd share is that for some of the cost validation tests that we've done, it's been remarkably consistent between two tests, the DxS assay, for example, and another quantitative PCR assay, that were used on the same specimens. Now, they may have suffered the same fate, we wouldn't know that, but to the extent possible, all the operational aspects of tissue storage were put in place.

DR. HARRINGTON: So I guess a quick follow up.

Have any of the sponsors who were speaking today looked hard at the consent issues to see whether they are solvable, whether those are insufficient explanation about why the material could be useful or whether there were language barriers or cultural barriers?

DR. YOUSSOUFIAN: So all of those are possibilities. One issue is that in some settings, in colon cancer, for example, in adjuvant trials, it's much more feasible to obtain tissue because they're coming off of surgery and tissue is available, for

example. In metastatic settings, that tissue availability may or may not be the same as in an adjuvant setting.

But, again, with all of the Erbitux trials and the panitumumab trials, as well, EGFR testing was an entry criterion. So by default, there was some tissue available and that was the basis for going back and analyzing them.

DR. DUTCHER: Dr. Reese, could you comment?

DR. REESE: Yes. I'd like to comment on this.

Our intent at Amgen -- and I could think one way to directly answer your question is really to make specimen collection mandatory. Our goal is to have a specimen available from every patient on every trial. If you look at the KRAS ascertainment rates across our studies, they're relatively high. They were 92 percent in the pivotal trial.

We have developed well defined internal processes for processing those specimens and then archiving them with the specific intent of obtaining or performing correlative biomarker analyses in the

future.

In our ongoing studies, the pivotal trials that I outlined to you today, we've successfully collected tumor samples from more than 98 percent of patients. We have developed specific consent forms that apply across all of our programs that permit collection and appropriate testing of these specimens.

(Pause)

DR. DUTCHER: All right. I'm going to go down the list of hands that were up.

Dr. Richardson?

DR. RICHARDSON: I have a question for Dr. Youssoufian. And I don't mean to be picky, but I'm just curious about the answer to this.

That is, in looking at the various data presented on the NCIC study, it talks about the demographic and baseline characteristics of subjects with the wild-type KRAS being similar to the overall population, except for those with ECOG Performance Status 2, which was lower in the cetuximab plus best supportive care arm compared to the best supportive care arm by itself. And it was a difference of

13 percent versus 26 percent, and this is in the group of patients with the wild-type KRAS.

I guess I'm curious what the breakdown is of Performance Status 2 patients in the KRAS wild-type arm and the mutation type, or putting it a different way, what do the overall survival data look like if you look at these PS-2 patients based on wild type and mutated status or if you exclude them and just look at the PS-0 and 1 patients, what do the numbers look like?

Do you have that information?

DR. YOUSSOUFIAN: Yes, I believe we do. So it actually turns out not to be different, but the difference has to do with the fact that at the time of the VGDS submission, we had incomplete data on these patients. And with the more complete information, that difference in ECOG Performance Status disappeared.

Let me have Nancy Gustafson, who is the statistician on this trial, and she could add further.

DR. GUSTAFSON: Nancy Gustafson, Bristol-Myers Squibb.

I don't have directly the results of an analysis which excluded the Performance Status 2

patients. However, the analysis that identified KRAS status as a predictive marker was done two different ways; one, without considering any prognostic factors or stratification factors, and, in the other case, adjusting for a wide variety of factors, including performance status.

You may recall that the interaction P value was 0.01 for the main analysis, without adjustment. It was 0.02 when performance status was taken into account. So that was how we addressed that imbalance.

DR. DUTCHER: Thank you. Dr. Netto?

DR. NETTO: Thank you. Just to go back to the issue of tissues. And not surprisingly, I'm a pathologist, so this weighs heavy on my mind. And we're going through the phase of getting orders on all of these KRAS tests. So several issues, I think, need to be addressed going forward and these --

DR. DUTCHER: These are questions for the presenters.

DR. NETTO: Correct.

DR. DUTCHER: Okay.

DR. NETTO: So the question is how can we

benefit from the existing material and kind of looking at certain questions that I'm sure from now on we're going to start doing studies on in terms of comparing different types of technology, looking at cutoffs, looking at micro dissection or not, looking at how much tumor is benign in the blood. There are several issues that will make a difference in terms of the standardization of the test that we end up facing with the HER2. So maybe we can benefit from all this retrospectively collected material in addressing this.

Are there plans for doing that, I think, because that's a source. Otherwise, we're going to have to await still the ongoing ordering now until we accumulate enough data and address this issue again and, meanwhile, maybe missing some patients, depriving them an opportunity of treatment.

DR. DUTCHER: Dr. Becker, do you want to take that on?

DR. BECKER: Well, I think that the issues are framed in terms of the part of the discussion that the questions that are posed for the panel to consider and the question, in part, that was asked earlier by

Dr. Mortimer.

We don't have, at this point, I think, an organized approach to being able to qualify specifically specimens coming out of deep archives, which is, I think, what you're describing. And so if you're looking at being able to go back and mine information that are even outside the clinical trial setting, that likely would be more problematic still than what we're trying to discuss here today specifically, where there might have been specimens either retrievable or set aside for the purpose of being able to look at how they would apply for the kinds of questions, predictive or prognostic, that we're dealing with here.

So I guess that the answer is that if one wants to go truly back to purely convenience samples that are pulled out of archives in pathology departments across the country, that is an even more problematic issue than the already significantly challenging issues that we're facing in the context that's being discussed explicitly today.

DR. DUTCHER: Dr. Curt?

DR. CURT: Thank you. It's a question for the agency and hopefully some comments from the sponsors.

I think this is a very important discussion today, because this retrospective/prospective scenario is likely to play out with biomarkers in the future and how we adapt to that, and working with the agency is important.

One of the questions I have is how does the agency validate or pass on a biomarker. Is it done, as is in the EU, by performance characteristics, where the label has already been changed or is there a need for clinical validation of the biomarker, in which case you might get into circular discussions with the sponsors on what the data really means in the absence of an FDA -- I don't want to say approved, but at least passed upon?

DR. BECKER: I'm not sure that I quite caught your question. I think that what you were asking is by what criteria do we end up allowing a pass on a biomarker. Okay, very good.

We do not deal with only the analytical

validity of the marker. There does need to be, as I showed off in, I think, the second slide, an indication of safety and effectiveness for the marker in some clinical context. And so the idea of a clinical validation for the biomarker is central to our being able to make a determination that the biomarker was appropriate for approval.

DR. KEEGAN: I think what you're asking is how -- in the scenario, for instance, that we've talked today, have we determined how they can do this given that there is no approved biomarker.

DR. CURT: Exactly.

DR. KEEGAN: And so the arrangement that we've reached is that the analytical qualification had to be acceptable to our colleagues in CDRH. And once they've reached that point, the analysis would both serve as looking at drug effects and would also serve to look at the predictive prognostic effects in the clinical trial, with something where we really had a good handle or a fairly good handle on the characteristics. And so it would proceed together.

DR. DUTCHER: Dr. Raghavan?

DR. RAGHAVAN: One of the important principles, as I spent my time on ODAC in the past, was that the mice are only allowed to guard some of the cheese. And so in the past, when applicants presented data to the FDA, they would present raw data and FDA biostatisticians would do their own independent analyses. That was always very important and often very entertaining.

I'm just wondering whether the FDA and, also, the applicants, and I don't mean specifically for this set of data, but generically, because that's part of what today is about, have considered this issue.

Will there be, ultimately, as biomarkers become more important, a mechanism to have an FDA-based bio repository? Will there be a requirement of the applicants to maintain a bio repository that is accessible? And are the applicants, potentially, at least, this duo of applicants, in a position to say that they would be able to provide tissue in the long term?

The technology changes regularly. The way we measure HER2/neu and many other things now is quite

different from what it was five years ago and the methodology is getting better. And so it may come to pass that rather than looking at discreet variables, the technology will allow us to look at continuous variables and the biostats will then change.

So tissue, will the FDA maintain it? Will you do your own assays? And how do the applicants feel about sharing tissue as a commodity?

DR. PAZDUR: Derek, I don't think the FDA will get into the business of doing a test and nor do I think that we will get into the business of being a repository for collection of all specimens. That would be a huge, huge undertaking.

We would require, obviously, looking at primary data. I think that is something we look at, but we don't, for example, serve as a repository for every EKG that was done in a clinical trial or every CT scan, nor do we review every CT scan that was done.

Here, again, we want to make sure that we have assurances that a proper analysis was done and the data is what it's purported to be. But for the agency to remain or be a repository for clinical specimens, I

don't think this is something that we have made plans to do.

DR. DUTCHER: Dr. Zhou?

DR. ZHOU: I have some technical questions about some analysis that's been done by the sponsor and, also, maybe a question to the FDA, too.

Can I ask multiple questions or just one question at a time?

DR. DUTCHER: If it's on the same topic, you can --

DR. ZHOU: Same statistic topics.

So there are three questions I want to ask about the data analysis presenters.

First, all the analysis done is the stratification for the KRAS, the strata is defined by the assay. As we know, the assay is imperfect. So they could have a misclassification of the people classified as KRAS positive -- wild type or the mutant.

So how do you adjust for the misclassification of the mutant KRAS and also the wild type? So that's the first question.

The second question is the first sponsor has

analyzed data using sensitivity analysis and, from my point of view, actually, the sensitivity analysis they have done actually is the single imputation. So basically, they imputed the missing assay results.

But they use a single imputation. In other words, they treated imputed data as real data. So that's why, in their P value, I think that's actually an inflated P value, because the standard deviation they calculate from the single imputation is much smaller than the two standard deviations. So the better way is probably to do the multiple imputation instead of single imputation.

The third question I have is about -- I think the second sponsor, they showed us a table of the treatment between the wild-type KRAS and the mutant KRAS, the mean number of infusion per patient. So for the wild type is like ten infusions and mutant is 4.9 infusions.

So that tells me, actually, you may have different treatment depending on the wild type. So that's kind of troubling me, because they should have the same treatment regarding of the wild types.

DR. DUTCHER: Dr. Reese?

DR. REESE: Dr. Reese from Amgen. If I might address the second part of your question first.

May I have a slide on?

So this shows patients receiving panitumumab with either wild-type KRAS tumors or mutant KRAS tumors. And as you note, patients with wild-type tumors received, on average, twice the number of infusions as those with KRAS mutant tumors. This is because patients with KRAS mutations progressed much more rapidly and were actually taken off therapy. So this really reflects the progression-free survival curves that we showed you and not an imbalance in actual treatments assigned.

Now, we tried to --

DR. DUTCHER: Excuse me, Dr. Reese. Just comment that infusions are weekly in both groups.

DR. REESE: That's correct. The infusions are weekly.

Does that adequately answer your question?

DR. ZHOU: Not really. If I'm understanding correctly, so the definition of treatment might be

different depending on the mutant group or the wild-type group.

Is that right?

DR. REESE: No.

DR. ZHOU: In terms of the chemotherapy part of it.

DR. REESE: No. That definition was not different. And one thing to consider is that this analysis was done retrospectively. Physicians and patients had no knowledge, nor did we, actually, of the KRAS status at the time treatment was occurring. Patients remained on therapy until the time of progressive disease. So the fact that patients with KRAS mutant tumors received fewer infusions is a reflection of the fact that they developed progressive disease more rapidly.

DR. ZHOU: So are there noncompliance issues, so that patients drop out because they're getting sicker in the trial?

DR. REESE: We have no evidence of noncompliance or patients dropping out for inappropriate reasons.

DR. DUTCHER: If you would, please.

DR. YOUSSOUFIAN: Hagop Youssoufian, ImClone. If I may address your second question, which had to do with the missing data scenarios that we presented.

That was only a subset of the data that our biostatistics group elected to show due to time constraints, and we, of course, chose the most extreme examples that we could think of; namely, that all of the patients came from one group or the other or that all the absent data came from one group or the other, and that's very likely not going to be a sort of real scenario. But even under that type of pressure testing, the interaction tests were strong.

So we have done additional analyses and, again, let me ask Dr. Michael Szarek from our biostatistics group to elaborate.

DR. SZAREK: Michael Szarek from ImClone.

So the scenarios that we've presented and what we've looked at in the NCIC study do represent single imputation. So we certainly could look at multiple imputation as a strategy, but we have not looked at that to date.

We have looked at additional scenarios where we assigned patients who are missing KRAS status to make KRAS more prognostic and we assigned them based on whether they had an event or not. So the patients who died during the study were assigned to one group or the other. And the predictive value of KRAS under those additional scenarios remained as measured by the interaction P value. But, again, with those additional scenarios, we could also potentially look at multiple imputations to see how sensitive the interaction P value is to that kind of strategy.

Also, referring back to your initial comment about misclassification, we have not attempted to adjust for potential misclassification based on the observed KRAS status data.

DR. DUTCHER: Thank you. Ms. Mason?

MS. MASON: My question is a bit more general, back to Dr. O'Neill or perhaps some of our other biostatistical folks.

How do you see this general discussion as impacting clinical trial design, like the numbers of participants needed for the intent to treat group? I

It seems to me it's going to have significant impact on that and representing consumers, that's a particular interest of mine.

DR. O'NEILL: If I understood your question, what impact does this have on the size of a clinical trial. Some folks will claim that if you get it right and you actually do this assay testing early enough that you become -- you have much more efficient clinical trials later on, because you're enriching the population for the likelihood of a treatment effect. So the downstream effect of early, good work that characterizes the sensitivity and the specificity, if you will, of the classification strategy should pay off in terms of a smaller trial.

I think the issue and the concern is smaller -- we've generally followed a practice of enriching trials for individuals. For example, in the cardiovascular area, you might take the more sick patients because you're likely to have more events. And so with a smaller sample size, you'd be able to find a treatment effect and then you might extrapolate later on to a less sick population or something like

that.

Here, the idea is, I think, that not only do you protect the negative group from toxicity, but you can get a more efficient trial, because the treatment effect is going to be larger than if you had an unselected population and it was sort of a diluted effect.

I think this goes to the issue of what's the impact of misclassification of marker status. Generally, equal misclassification in the treated and the control group should drive the interaction effect more to a null or a zero effect. So it would be harder to detect a difference between the positive and the negative group.

But I think others have written about this. I think Dr. Simon has written extensively on the efficiency of clinical trials. That is very much a function of the properties of the screening strategies.

DR. DUTCHER: Dr. Pazdur?

DR. PAZDUR: To answer your question from a different perspective, and I'm not referring to these companies and these products at all, but I think

inherent in this whole process is a thoughtfully planned out process and doing your homework before you enter into a Phase III study. And, unfortunately, that isn't the case. Many times, we find this rush, rush, rush to a Phase III trial, many times, sponsors even wanting to skip a Phase II trial because they've seen a few responses in a Phase I trial.

So this whole thing is predicated upon a very thoughtful deliberation, and if one was going to do this in a very prospective fashion, one would want to have a clear understanding of the mechanism of action of the drug, development with a partner, for the in vitro diagnostic. This is one of my worries and one of the worries that promulgated this discussion here is the emphasis, yes, we can go backwards, but there are problems with going backwards, and if people are planning it, they need to plan in a prospective fashion with some thought being given.

DR. DUTCHER: Dr. Simon?

DR. SIMON: I'd like to just try to also respond to your question.

I think in the perfect situation, say, the

Herceptin situation, it leads to a smaller randomized clinical trial because the patients who get into the trial actually benefit, whereas you don't have a lot of patients being randomized who actually don't benefit; also, you have a set of patients who you know who benefits.

But I think in real life, developing drugs with predictive biomarkers will actually lead to larger clinical trials and it's going to make life not simpler, quicker and cheaper; it's going to make life more complicated, more costly, and it's going to require larger clinical trials because you're rarely going to have -- often, you're not going to have full confidence in your diagnostic by the time you get to your pivotal trial. You're going to wind up having to have enough test positive patients to analyze them separately and test negative patients to analyze them separately.

Even in cases like the Herceptin case, where you think you really know who is likely to benefit from this drug, based on what Dr. Becker was pointing out before, the diagnostics, CDRH, they want you to include

test negative patients anyway. And I think that creates serious issues for patients. You wind up having to, if you're honest with patients, say "We have this drug that we don't think is going to help you, but in order to show the FDA that it doesn't help you, we would like you to participate in this trial."

DR. DUTCHER: Dr. Harrington?

DR. HARRINGTON: I wanted to second what Dr. Simon is saying from one more perspective.

I think that what we're seeing here is one example of something we will see in the future quite a lot, which is where the science races ahead of the trials. Trials are notoriously difficult to plan. They're planned well in advance and the markers sometimes become available later.

So I agree with Dr. Simon that the trials and practice will become larger because we will be aware that there are likely heterogeneous subgroups in these trial populations, but may not quite know how to identify them when the trials get started.

So I think that not only will they become larger, but I think there will be increased

requirements to make sure that those tissues are available and archived in a viable way so that these kinds of retrospective analyses can be done with more confidence.

DR. DUTCHER: Doctor -- I can't say your name, sorry.

DR. PRZYGODZKI: Przygodzki.

DR. DUTCHER: Przygodzki.

DR. PRZYGODZKI: I have a couple of questions to the sponsors regarding the tissues and the nuts and bolts of how the testing was actually performed on them.

An earlier question was brought up by, I guess, Dr. Mortimer, regarding stability of DNA and all that and how RAS can mutate or not mutate. In the early part of the carcinogenesis, you may or may not have RAS that's wild type. But more often than not, actually, in the early stages, like I mentioned, the invasive front may mutate and actually that will be the greater population that, at one point, will overcome the entire tumor in and of itself with the mutation. In this sense, the metastatic tumor would be likewise

mutated.

In this situation that we're looking at right now, that whole issue really isn't that big of a problem vis-a-vis the idea of taking a biopsy of the tumor from the colon versus having to explant itself and then studying it. But along that train of thought, there is the question of truly looking at the metastatic versus the primary in the sense that -- and these are essentially the questions to the sponsors.

Is there truly histopathologic evidence of the sample that you have truly showing tumor? The reason I bring it up is if it is in a lymph node, you may be sampling normal cells and you may be getting wild type. And on top of that, you may be also looking at a sample that just shows lymphocytes in a theoretically positive tumor, and positive lymph node, that may be, at one point, positive, but as you're trimming through the sample, you may wind up with something that's negative. And that's, I guess, one point.

But the other thing is, also, along the line of what is your cutoff for RAS positive, because as I

recall, when you look at sequence data, you're looking at around 10 to 15 to 20 percent of the cells in the test to be truly positive to really be called as positive, as mutated. But with the assays that are being used now, the cutoff is getting lower and lower and lower.

So I wonder, is it two percent of the cells positive considered as being positive, that is, mutated and, therefore, this person is off the chart, where, in de facto, may be, under all considerations, wild type? I just wonder.

And one final question. Samples, were they taken pre-therapy of any kind? Is it the biopsy itself or is it after therapy, as well? Is it a mix or what?

DR. REESE: Dr. Reese from Amgen.

To address your questions, if I can remember them all, most of the samples assessed were paraffin-embedded specimens that had been obtained at the time of primary resection of the primary tumor, the vast majority of those.

One of your other questions related to ensuring that you actually have tumor present in the

sample. As part of our sample processing, these samples go through a rigorous quality assurance and, in fact, a pathologist must confirm that tumor is present, marks that tumor on that slides, from which DNA will be extracted, so that we're certain that we're getting tumor DNA.

Finally, if I might have a slide on, you asked a question, I think, that relates to the sensitivity. And you raise, actually, a subtle, but important point about perhaps what the gold standards ought to be when we're developing some of these new tests.

The DxS KRAS mutation test kit that we use, as you can see in the second bullet, can detect approximately one percent of mutant KRAS DNA in a background of wild-type genomic DNA. It requires about five or ten copies of mutant DNA to read out as positive.

Now, what's interesting is that this is more sensitive than typical sequencing, where you, of course, have add mixture of normal cells, stromal elements and other things. And, as you have pointed

out, you often require 15 to 20 percent tumor to get an accurate result.

I think this raises a broader question for the committee. As science is moving in real time and new assays are being developed that may be actually superior to existing gold standards, how do we qualify those assays? How do we define sensitivities, specificities, precision, and all of the other variables we care about?

DR. PRZYGODZKI: So what was their cutoff for a tumor to be positive, what percentage?

DR. REESE: Let me have Dr. Scott Patterson of Amgen Medical Sciences, who has done extensive work in this area, comment on that.

DR. PATTERSON: Thank you, Dr. Reese, and thanks for the question.

We have currently in our qualification process been examining the smallest amount of tumor material that can be required. If one is to conduct the testing without any trimming of the section, then we require 20 percent of that section to be of tumor. However, we can go down to much smaller amounts of

tumor, determined by the pathologist, and then remove some of the normal tissue.

So you can actually get down to significantly less than four square millimeters. In fact, we're continuing to examine that down to potentially one square millimeter.

DR. PRZYGODZKI: That's perfectly okay. The question, though, is within the tumor, what percentage of actually tumor cells are RAS mutated, that you call as RAS mutated? One percent, five percent?

DR. PATTERSON: That's a very interesting question. We haven't been able to determine on a per cell basis the KRAS mutation status. There are actually no tools to do that. And I think that that's a very interesting scientific question that we'd like to further pursue.

DR. PRZYGODZKI: I understand. But you're doing this fluorescently with a CT level and you are gauging that, also, toward a certain copy number.

Would it be possible to extrapolate what percentage of those cells?

DR. PATTERSON: The sensitivity of the assay,

as we had, if we could have slide on that we previously showed, the work that DxS has done has shown that the sensitivity of the assay is down to five to ten copies of mutant KRAS.

DR. PRZYGODZKI: Okay. Thank you.

DR. DUTCHER: Dr. Grem?

DR. GREM: For the sponsors, there were variable, maybe up to 90 percent, 92 percent acquisition.

Now, were all of those samples informative? Were there any samples that you just didn't get and the reaction didn't work and how were those patients handled then?

DR. REESE: David Reese from Amgen. Slide on.

This is a slide that we showed you from the core presentation. Tumor samples were available from 96 percent of all patients in the trial. KRAS testing failed in four percent of those cases uniformly due to either insufficient DNA quality, i.e., the DNA was degraded for whatever reasons, or inadequate quantity of DNA.

DR. GREM: So were those patients who had a non-informative sample, were they analyzed or just excluded? I mean, what do you do in the clinical setting if you send off a tumor specimen that just had not enough tissue available? Do we deny those patients therapy or re-biopsy them?

DR. REESE: Right. I think that's a very good question. In our view, optimally, one would obtain additional tissue and attempt to ascertain KRAS status. In the absence of that, the current label for panitumumab, of course, is in all comers, so physicians administering the drug to patients in that setting would be prescribing according to the current U.S. label.

DR. DUTCHER: Dr. Youssoufian?

DR. YOUSSOUFIAN: Youssoufian, ImClone.

Just to add, briefly, the ImClone/Bristol-Myers Squibb experience. So far, we've had that opportunity to do correlations in two different studies, EPIC and OPUS. In EPIC, the KRAS evaluable population that we described, approximately, out of 60 samples or so that were tested, 98 percent, a

very high number, were informative of KRAS analysis. In OPUS, it was in the low 90s.

DR. BECKER: I just would like to echo a little bit of what has been asked by Dr. Przygodzki and I think, to a lesser extent, by Dr. Grem concerning the impact of understanding what the assay is telling you concerning the status of the cells or the tissue as a whole, that these do bear then on what is the performance of a cut point that's used to divide patients between marker positive, who would be considered likely to benefit from a drug, for example, and marker negative, who would likely not benefit.

From the perspective of CDRH, we certainly are not interested in seeing trials accrue patients for which there is a settled knowledge that they cannot benefit from the drug on the basis of, one hopes, a well validated biomarker.

However, to the extent that that has not been settled definitively, then the opportunity of being able to get to a most expeditious demonstration of the absence of benefit in the marker negative patients is one that we're interested in hearing ideas about how to

be able to achieve that as quickly as possible.

The idea at the end of the game then is to be confident that when that patient obtains a negative test for the marker as deployed and says, "Gee, Doctor, why is it that I can't take this drug," that the answer isn't that "It's our best guess that you won't benefit."

DR. DUTCHER: Dr. Link?

DR. LINK: My question sort of relates to that, too. In terms of the cutoff point and what you lose in having an all or nothing kind of cutoff is that you lose the quantitation and the range. So in other words, it may be very different if somebody is 100 percent positive for mutated versus at the level of detection, and we find that in other diseases, as well.

So is there any attempt to try to make this a quantitative assay and actually relook at this data? Unfortunately, there's not that much benefit in the patients who benefit. But in tumors where there's a lot of benefit potential from the therapy, it would make a big difference if you could change your cutoff or you could give some quantitation of the likelihood

of benefit.

DR. YOUSSOUFIAN: Youssoufian from ImClone.

Actually, we do have an expert who is very much involved with the DxS assay development, and this applies to both sponsors. So maybe Dr. Little would like to comment.

DR. LITTLE: Thanks. Hello. My name is Steve Little. I'm the CEO of DxS.

DxS is a U.K. personalized medicine company. We develop biomarkers and companion diagnostics to help predict response to drugs. So we're delighted to be working with both Amgen and ImClone on these projects.

What I'd say about the quantitative aspects of this particular test is that we have not attempted to make the test quantitative. It seems that a lower level of mutation or a high level mutation both appear to correlate with a lack of response to the drugs. But perhaps it may be that the sponsors would be in a better position to talk about mutation percentage and actual drug response.

DR. LINK: In this particular clinical setting, most patients don't respond. So it might be

more helpful that the patients who weren't responders were 90 percent positive as opposed to one percent positive.

DR. REESE: So if I just might add to the discussion. I think one way to frame it may be that if using the DxS test, we classify a patient as KRAS mutant, the negative predictive value for response at least is 100 percent. Across our studies, in more than 300 patients with KRAS mutant tumors, none who was classified as such has a response. Now, that, of course, doesn't speak to the KRAS wild-type group and, of course, we need to, obviously, make further improvement there.

DR. DUTCHER: I think we need to move on to the open public hearing and then for those of you that have other questions, we can address them with the sponsor and the FDA in the afternoon when we're going through more general discussion.

Nicole?

MS. VESELY: Both the Food and Drug Administration and the public believe in a transparent process for information-gathering and decision-making.

To ensure such transparency at the open public hearing session of the advisory committee meeting, the FDA believes that it is important to understand the context of an individual's presentation.

For this reason, FDA encourages you, the open public hearing speaker, at the beginning of your written or oral statement, to advise the committee of any financial relationships that you may have with the sponsor, its product, and, if known, its direct competitors. For example, this financial information may include the sponsor's payment of your travel, lodging or other expenses in connection with your attendance at the meeting.

Likewise, FDA encourages you, at the beginning of your statement, to advise the committee if you do not have any such financial relationships. If you choose not to address this issue of financial relationships at the beginning of your statement, it will not preclude you from speaking.

The FDA and this committee place great importance in the open public hearing process. The insights and comments provided can help the agency and

this committee in their consideration of the issues before them.

That said, in many instances and for many topics, there will be a variety of opinions. One of our goals today is for this open public hearing to be conducted in a fair and open way, where every participant is listened to carefully and treated with dignity, courtesy and respect. Therefore, please speak only when recognized by the chair.

Thank you for your cooperation.

DR. DUTCHER: So we have three people who have asked to speak at the open public hearing. They will each have five minutes. We ask that they please identify themselves and any connections with the sponsors.

MR. ERWIN: I'm Robert Erwin, President of the Marti Nelson Cancer Foundation. Neither the foundation nor I personally stand to gain or lose financially from today's discussion. We have received no contributions or sponsorship from any company in the biotech, pharmaceutical or medical device fields for the past four years. We're an all volunteer, nonprofit

organization, provide all of our services free of charge, and I bought my own airplane ticket to come here.

The unfortunate reality is that cancer drugs and biologics do not work very well or for very long for many patients and a large percentage of people who are treated with them receive no measurable benefit. However, virtually all treated patients are harmed in some way, even if only temporarily. From the standpoint of an individual making a choice, it comes down to the risk of harm with the possibility of benefit versus the risk of harm with no possibility of benefit.

In the current case, with the example of KRAS, we're dealing with a marker that is predictive of non-benefit and I think that that semantic reversal from predictive of treatment benefit to predictive of no benefit is an important consideration in this particular case.

It's not easy to simultaneously develop targeted drugs and diagnostic products to guide their use, particularly since they're usually developed by

different companies and coordination can be difficult.

The processes and standards for proving effectiveness of the combination is complicated. To further complicate matters, there are companies very willing to foist products on the market that have not been proven to be of value and this has been true for both some attempted therapeutic products and diagnostic products.

It's probably not practical for the FDA to regulate all laboratory tests because of the overwhelming workload that would create, which is not helped by the government's chronic unwillingness to properly fund the FDA.

So your discussions today are really important from the standpoint of providing greater clarity for future situations beyond KRAS, to encourage the development of these products, while maintaining very high standards for approval, keeping in mind the ideal of what is optimal, but sometimes acting on the basis of what is pragmatic.

This is essentially a safety issue, in our view, and limiting the labeled indications of the EGFR

therapeutic products to exclude treatment of KRAS mutant patients is rational and important for patient safety, at least in colorectal cancer treatment outside of the context of clinical trials.

Safety should be regarded not only as avoidance of toxicity in the absence of therapeutic benefit, but, also, the avoidance of losing time with ineffective treatments when every day or week counts. Wasting time with a treatment that does not work may shorten life by delaying access to a treatment that does work or it may make a person ineligible for a promising clinical trial.

The fact that the most measurable benefit of these particular agents is PFS rather than overall survival does not change the fact that the KRAS analysis predicts an absolute lack of benefit.

It's beyond the scope of this meeting to discuss the clinical value of PFS, but we're one organization that does value it. We believe that good practical judgment suggests that the prospective/retrospective analyses that have been discussed and presented today can legitimately be used

to guide the use of EGFR inhibitors. We do not believe that this would open the floodgates to unscrupulous people who would love to use invalid statistical dredging techniques to get worthless products approved.

We are pleased that the sponsors are actually advocating restrictions in their own products labels on the basis of science and public policy and we applaud the FDA's difficult work to strike the right balance between acting on evidence sufficient to guide the current practice of medicine, while maintaining high standards for future products.

We hope your discussion and debate will be practical today with respect to what we know both about this specific case and with respect to helping to clarify the path that will bring legitimate, successful future innovations to the clinic faster and to the patients who need it.

Thank you.

DR. DUTCHER: Thank you very much.

The next speaker is Carlea Bauman.

MS. BAUMAN: Hello. I'm Carlea Bauman. I'm the President of C3 Colorectal Cancer Coalition. C3 is

a nonprofit, nonpartisan advocacy organization committed to winning the fight against colon and rectal cancer through research, empowerment and access.

C3 receives funding from both Bristol-Myers Squibb, ImClone Systems and Amgen in the form of charitable donations. Additionally, in 2008, C3 received a charitable grant from Caris Diagnostics, a company that tests colorectal cancer tumors for the KRAS mutation.

None of these companies or any of our other corporate supporters has influenced our comments on this issue.

On page 3 of the FDA briefing document, FDA listed the requirements they laid for submission of KRAS data. In theory, these are good requirements and could be applied to both KRAS and future biomarkers and companion diagnostics. However, we feel that the requirements fail to take into account the fact that there are many approved drugs on the market. As we speak, researchers are mining banks' tissues looking for markers that will help target those drugs most effectively.

Some of that research may yield gems that will help patients avoid toxicity or increase likelihood of benefit from treatment. However, while current and ongoing trials routinely bank high proportions of tissues, past trials did not. Thus, a requirement that tissue be available from 90 to 95 percent of participants in a single trial means that tissue from past trials will be largely useless for submissions.

In this era of personalized medicine, C3 strongly believes that we must be cautious about one-size-fits-all requirements. We believe that tissue and assay requirements vary and that the following types of questions should be considered.

What do we know about the biomarker? Are the results consistent with what we know about the mechanism of action? Do we know what percentage of patients in the general population have this biomarker?

What do we know about the assay? Does the assay involve new technology? Are the assay results subjective, such as in a gene expression test, or are they objective, such as in a gene mutation test?

Do they require black box calculations, such as Oncotype DX? Are the clinical results consistent across analyses of multiple datasets? What is the strength of the clinical impact on patients? Do patients with a specific biomarker respond more to a specific drug, less, or not at all?

With respect to KRAS, getting hold of 95 percent of the tissue across the spectrum of trials is not possible. However, the tissues that have been analyzed show a consistent breakdown of about two to one between wild type and mutant patients, which is consistent with the occurrence of the historical mutation rate in colorectal cancer.

The suspected mechanism of action is consistent with lack of response in the KRAS mutant population. The results hold true across trials, regardless of phase, assay or laboratory. The assay results are black-and-white and gene mutation tests are not a new technology. KRAS mutant patients get no benefit from cetuximab or panitumumab. This has been shown in multiple analyses. And in OPUS, KRAS mutant patients who received cetuximab did worse.

With respect to the KRAS issue, we have reviewed the analyses, spoken with researchers and looked at positions taken by the National Comprehensive Cancer Network and the College of American Pathology. As a result, we strongly urge ODAC to recommend a label modification for KRAS to provide some mention of these research findings in the label, perhaps something as simple as research indicates that patients with KRAS mutant tumors do not benefit from treatment with this class of drugs. In addition, the KRAS story illustrates the problem with one-size-fits-all requirements for companion diagnostics.

We urge ODAC to recommend that, for the future, FDA develop requirements fit with the reality of the research environment and the wide variability of markers and assays. At the end of the day, a flexible approach will facilitate the development of assays that allow patients to benefit from the reality of personalized medicine.

Thank you.

DR. DUTCHER: Thank you. And our third speaker is David Apelian.

DR. APELIAN: Thank you. My name is David Apelian. I'm the chief medical officer at GlobeImmune, a clinical development stage therapeutic vaccine company based in Boulder, Colorado. I want to thank the committee for allowing me a few moments to share some of our observations from our clinical testing of RAS in our patient populations.

In contrast to the discussion of this morning, our motivation for testing patients is to proactively characterize their RAS mutations so that we can specifically target our therapeutic vaccine to the exact RAS mutation in their tumor. So this is an entry criterion for patients to be eligible for treatment with our particular RAS targeted vaccine therapy.

The committee is probably very familiar with this representation of RAS, but I show it to you to represent that we are continuing to learn about the key mutations that activate RAS, and it does point to the fact that we probably are just beginning to understand the complex nature of this characterization.

Of course, the position 12 mutations and Exon 2 still constitute the majority of the activating

mutations in KRAS, particularly G12 to V, C or D. But we know now, from our own data and public databases, that less than typical mutations in Exon 2, as well as Exon 3, actually comprise about 25 percent of the activating mutations in RAS.

So this does represent a fairly large subset of the patients that, if we're not testing for those mutations that, on an individual basis, are not that frequent, but collectively contribute in a meaningful way to the overall proportion of mutated RAS, we could have an invisible false negative artifact in the way we run these assays and identify patients for eligibility for treatment.

I think another important point is that even in Exon 3, we've discovered a novel mutation at position 76, which occurs in tandem to the more commonly found mutations in Exon 2, namely, at position 12 in most cases. And we find this tandem mutation more commonly in advanced cancer patients or metastatic patients and, in fact, our in vitro testing of this double hit in RAS predicts a more aggressive invasive behavior in vitro based on auger invasive endpoints in

our in vitro testing, which further indicates that not only are some of these novel mutations informative in terms of activating RAS, but may interact between Exon 3 and 2 and teach us even more about how this marker could behave and predict response to therapy.

So I would point out that perhaps it's over-simplified to say that someone is either a wild-type RAS or mutant RAS patient and we should perhaps focus more on fully characterizing the mutations in a particular patient sample to fully understand the effect on therapy.

In our hands, though, we still consider the gold standard optimized bidirectional sequencing and we find this to give us the best specificity and sensitivity in our clinical samples. We've used a published procedure called PNA clamping to reduce the wild-type signal, which partially addresses the fact that we often get samples with normal cells mixed in and even tumor cells will have a wild-type allele, which will present a signal in these types of assays. So we've used this combination approach of a squelching mechanism with bidirectional sequencing to improve our

signal-to-noise ratio.

While I think the commercially available assays are a significant advance in terms of the convenience and efficiency of testing, in our hands, at least, we can't get that same high level of specificity and sensitivity using those commercially available assays.

I think, again, we can only find mutations if we look for them. So if we exclude Exon 3 or exclude less than typical Exon 2 mutations, we're going to have an artifactually high but invisible false negative rate on these patients, with serious implications about now allowing patients to receive therapy with very little chance for benefit.

We've sequenced 415 clinical samples in our programs. This includes archived tissue sets, Phase I patients with pancreas, colorectal and non-small cell lung cancer, as well as almost 200 patients in our Phase II study in newly diagnosed resected pancreas cancer. And we've seen the ability to improve our response rates using the optimized bidirectional sequencing by about 15 percent in the pancreas cancer

patients. So we're fairly confident that this is at least currently the best standard for us to use for qualifying patients for our randomized Phase II trial in resected pancreas cancer.

Just to illustrate, in the rightward histogram here, in fact, the classic mutations, G12 to V, C and D, do comprise about 75 percent of our Phase II pancreas cancer mutations, but 25 percent of those mutations that we've identified are atypical Exon 2 mutations or Exon 3 mutations. And this is consistent with the middle bar shown here, which is our Phase I population, as well as the publicly available database called COSMIC, which, again, shows that the atypical Exon 2 mutations and Exon 3 mutations actually make up about 25 percent of the samples tested. So it does point to the need to make sure we're addressing a full complement of these mutations before we call someone a negative KRAS patient.

In conclusion, we think RAS does present incredible promise as a predictor for response to therapy, but the implications of having a false negative rate anywhere from five to 25 percent does

have serious implications in terms of treating patients with very little opportunity for benefit when we consider the response to EGFR targeted therapies, whether they be monoclonal antibodies or even the TKI class, as we've seen in lung cancer, showing poor response in RAS positive patients.

So I think one of the questions I'll pose rhetorically to the committee is how low is low enough for false negative rate or false positive rate with these assays? Is five percent good enough? Is one percent good enough?

I think when you frame that in the context of the implications for the patients' response to therapy in terms of the risk-benefit to patients and the cost benefit to the third-party payers, this is still a very important question.

Thank you.

DR. DUTCHER: Thank you very much.

That concludes the open public hearing. We are going to take a break for lunch. We will reconvene in this room in one hour from now. So it's 12:07, so I guess 1:07, say 1:10.

Please take any personal belongings you may want with you at this time. Committee members, please remember, there should be no discussion of the meeting during lunch among yourselves, with the press or with any member of the audience.

Thank you.

(Whereupon, a lunch recess was taken at 12:07 p.m.)

A F T E R N O O N  S E S S I O N

DR. DUTCHER: Okay. We're going to start the afternoon session. Today's discussion focuses on the type and amount of data needed to support product labeling using biomarkers.

In the following discussions, we are assuming that prospective studies intended to establish the clinical usefulness of the biomarker have not been performed and that decisions are being requested that require a retrospective analysis of a completed or ongoing clinical trial.

For the following series of questions, assume that appropriate tumor sample acquisition and handling procedures were used, the assay for the biomarker has acceptable analytical validation, and clinical data would be obtained from randomized controlled clinical trials. This discussion applies to studies which met the pre-specified primary study endpoints and would not be intended as a mechanism to salvage failed trials.

So the way we're going to approach this, the FDA has given us five questions and has asked members of the committee to serve as discussants to start off a

conversation. So we're asking that the people that are discussing limit their comments to three to five minutes, so that we have about 20 to 25 minutes per question for the rest of the committee to discuss.

We'll take comments by show of hands and I apologize if I cut you off in advance. It's nothing personal. But we want to give as many people a chance to say what they have to say as possible.

So that's going to be the format. And somebody told me it was going to be like hurting cats. So I like cats.

All right. We're going to start with question number five, because one of our discussants has to leave a bit early.

Please discuss the importance of timing and rigor in determining the analytic performance of the companion diagnostic test.

Who would like to begin?

DR. ZHOU: Okay. So I will start out with some questions I have about how to evaluate the discrimination performance of the biomarkers.

So the first question I have is what type of

scale we should use for the biomarker. Right now, we all heard the result from the biomarker is binary, either mutant or non-mutant. But I think the original scale, actually, probably is a continuous scale. So if it's a continuous scale, how does it go from the continuous scale to the binary scale? So that's really the issue about how do you choose a cutoff point.

Should cutoff point issue depend on the particular outcomes that were used? Because right now, from the discussion we had in the morning, we have three types of outcomes, the progression disease free or the total survival or the response.

DR. DUTCHER: Excuse me. Dr. Zhou, this question is with respect to the biomarker diagnostic test and not to have anything to do with what we talked about in the morning in terms of the actual studies.

DR. ZHOU: Okay, sorry.

DR. DUTCHER: So this is supposed to be a general discussion.

DR. ZHOU: Yes, it is general. I think the four other biomarker studies also had sort of those kind of different responses.

So the question is, should the cutoff point we choose for defining either marker positive or marker negative depend on the patient characteristics and also depend on particular outcomes?

The second question is, how do we measure the diagnostic accuracy of the biomarker when the outcome is timed to the event?

So most of the literature we see actually suggests we use sensitivity and specificity and RC curve. But the problem with sensitivity and specificity and the RC curve is that they require the true disease status to be binary. It should not depend on the time. But, however, if the outcome is timed to the event, actually, the disease status of the patient changes depending on what time you look at. If you look at six months or the one year, the disease of the patient actually might be different.

So better measurement for measuring the discrimination capability of biomarkers maybe should use the so-called time-dependent RC curve or time-dependent sensitivity and specificity. I have seen that in a lot of the literature when they try to

evaluate accuracy of diagnostics. I think this issue should be considered by the FDA and, also, the sponsors.

The next question is about gold standards. So what is the gold standard we should use when we try to evaluate accuracy of diagnostic biomarkers, because it's not very clear at this point.

What is a true gold standard, for example, for the mutant, the KRAS, for example?

The next issue is, what's the impact of imperfect assays on the evaluation of either prognostic or predictive biomarker values?

The last one is, should we only rely on key value when we try to establish the predictive value of the biomarker? Because the P value is only for the known hypothesis. They say nothing about alternatives.

So in other words, what is effective size the predictive marker should be? It should be like median survival, five-year versus two-year, or should it be the hazard rate, two or three?

So those kind of issues I think should be clarified in our discussion in the committee.

DR. PRZYGODZKI: I guess from the standpoint -- I agree with Dr. Zhou's mentioning it. I just wanted to add the point of, okay, so how does one truly go about, from the histopathological end, that is, accruing of the tissue.

Ultimately, this is the real meat of the whole situation. With the result of whatever happens here, everything else goes down the path depending on if it's positive, mutated, that is, or wild-type negative, the tumor, that is.

Issues that come to mind are, as I mentioned, I guess, earlier, when does one actually accrue the sample? Is this as from the resection specimen itself, from the initial diagnostic biopsy that this was performed on? Should one look at the lymph node itself, the distant metastasis itself, the primary tumor? What is enough in sample size to actually diagnose?

Current standards, one could look at a small biopsy and get enough information and do a decent study on that. Yet, if one looks at this going off to the other extreme, where everybody in the world is really

beginning to use this in this method, we have to have at least some guidelines of what is the typical approach that one has to use to really make this as even of a type of test, a type of test in the sense of accrual of tissue and what we use for the test itself, relatively standardized.

To boot, on top of this, the level -- echoing, again, what Dr. Zhou was mentioning and mentioned earlier, as well, what is the true cutoff for positivity?

Now, if we, again, go back toward the idea of the gold standard that we use currently, which is bidirectional sequencing, we're looking at 10 to 20 percent of tumor cells that are mutated to be diagnosed as mutated.

In the current pathway that the folks are using now with the DxS method, the sample, of course, is much more refined and one could really identify a percentage, a small percentage of cells that are mutated.

The question is, is a sample that's five percent mutated truly entirely mutated? If one uses

the gold standard of sequencing of 20 percent as being the gold standard, and, there, you have 20 percent that is actually mutated, I think we need to have a clear cutoff to at least establish what is truly mutated or not in said samples.

My five cents' worth. Thank you.

DR. DUTCHER: Okay. So we can open this question up to discussion. Let me ask a question of the pathologists.

Are there any -- and I think the answer is probably no. But are there any guidelines for sample collection in clinical pathology and block construction that would be representative of the specimen or is it dependent on who is doing the cutting and who is doing the looking?

DR. PRZYGODZKI: There are general guidelines. You take margins of the resection to show that there is no tumor. You take the tumor itself and relative idea of one sample per centimeter. Usually, people take more than that.

Essentially, all of the lymph nodes, there are pericolic lymph nodes that are accrued in your

samples, as well, that are taken and split and put into cassettes.

It isn't truly different from place to place. Essentially, if you misdiagnose or under-diagnose, it is criminal. One needs to make appropriate diagnostics on that sample. So one does go the extra mile to truly make the diagnosis as accurate as possible. So in that sense, it's pretty standardized.

DR. DUTCHER: And what does the clinical trial get when someone asks for blocks?

DR. NETTO: So basically representative of the tumor. But the broader question is, is it the primary or is the metastasis and which part of the tumor, the cutting edge of the tumor, interface with the benign? Do you need to micro-dissect or not? Do you need to circle for the lab who is extracting the DNA or not?

So there is no standardization for that. But generally, you try to pick the block with most tumor-to-normal ratio in order to increase the chance.

DR. PRZYGODZKI: Actually, some molecular diagnostics labs will go into triplicate sampling. So

in other words, you will take either three different sample areas or three different blocks and try and see if the alteration is similar.

Granted, that's a lot more money that one expends to make the diagnosis as it is and it depends on what type of molecular test that one is looking at.

In this case -- usually, in my previous life, when I was actually doing this in the lab, one would take at least two or three samples to really accrue and make that determination, at least bidirectional.

DR. DUTCHER: Dr. Wilson?

DR. WILSON: To just get back to this specific question, I think we all would agree that it's optimal to do this in a prospective manner.

But I think the reality is, number one, many targeted agents, or many putative targeted agents, one finds out over time, are not truly targeted. So any biomarker you have may go from a mechanistic-based biomarker to simply a prognostic biomarker and, therefore, may affect both patients that have a mutation or not, such as we have seen earlier.

But I just wanted to get back to the example

at hand. And I think that one of the pitfalls, for example, besides these statistical ones, in terms of doing the current study after the fact is that among those patients that are EGFR negative by IHC, we don't know whether or not a wild-type RAS might, in fact, identify a patient who might benefit from a drug.

So I do think that missed opportunities come about, as well, by doing this later as we begin to understand the biology. But I think at the end of the day, we can talk about the importance, but this is going to be driven by the availability of these types of studies, of these types of tests, and, often, they're just simply not going to be available when we're doing the initial upfront studies.

DR. DUTCHER: Dr. Harrington? Dr. Link?

DR. LINK: I have two questions. I don't know if this is the right question to address.

Number one is sort of the -- they have to do with evolution. The tumor evolves and so does the testing evolves.

So the question is would a patient be eligible for a trial who had a biopsy on hand that had

been done of his primary tumor now that he comes in for a study of metastatic disease.

So what's going on -- I mean, certainly, we know from leukemia studies that the leukemic clone at relapse may be very different in terms of mutations.

So I think you're going to have to specify that the tumor is going to have to be in time proximal to what you're doing, and I don't even know if we know that from the others, because some of them said that they were from primary tumor resections of this, whereas the patient had metastatic disease and some of these were patients who had developed recurrence. So it's not clear that the tumor specimen represented what the patient looks like now at the time they're being treated.

The second thing is -- I don't know where to bring this up, so now better than -- because I may not get the floor again. But testing evolves. So we have a sensitivity that has been down to one percent now. There will be methods, I'm sure, over a relatively short range of time, if people are really interested in finding out the answer, that will be more sensitive.

So the question is, how do you know when that test is or are you going to have to run this whole thing over again when you have a more sensitive test that maybe gives you a sensitivity down to another order of magnitude?

Are you going to have to go through and troll through this again to find out if you can identify, even further refine it or slice the bologna even thinner in terms of really refining who is going to be a responder and who won't?

DR. DUTCHER: Dr. Simon?

DR. SIMON: Well, I think this is talking about analytical validation and so reproducibility and does it agree with the gold standard, if there's a gold standard.

But to follow-up on what Dr. Link was saying, I think the sort of mindset -- I mean, I think, ideally, you want to have an analytically validated test used prospectively in your pivotal trial. But because of the complexity and because science often is out of synch with clinical development, I think very often that is not going to be the case. And if we're