FDA-AACR Immuno-Oncology Drug Development Workshop
Transcript from October 14, 2016

Section 1 of 45 [00:00:00 - 00:10:04] *(NOTE: speaker names may be different in each section)*

Suzanne Topalian: Good morning, everybody. Good morning and welcome to day two of the IO drug development workshop co-hosted by FDA and AACR. Yesterday we had a pretty intense day of very interesting presentations and panel discussions. We started by
[00:00:30] hearing about the principles of immuno-biology, underlying the practice of cancer immunotherapy and there we talked about checkpoint inhibitors as well as vaccines and ACT. Then we had a session about the pre-clinical evaluation of IO products. We heard about novel mouse models and vitro assays and also the study of bio-markers in human tissues. We heard about considerations for dose findings
[00:01:00] for new IO agents, including agents used as mono-therapies or as combinations. We wrapped up the day talking about how to evaluate immune-related adverse events.

Today, we're going to start by having two talks from the FDA about traditional and alternative endpoints for IO development and then we'll have three sessions. The
[00:01:30] first will focus on traditional endpoints and how they might apply to IO development. The second will be about alternative endpoints. Then we'll use all of that knowledge about selection of the right endpoints to see how that would play into novel clinical trial design. We're going to wrap up the conference with a session about novel trial design, including trials for combination therapies, bio-marker-driven trials, adjuvant, and late-stage development trials.

I would like to remind everybody to please use the microphones, including the
[00:02:00] panelists and people asking questions, because we do have over 1,000 people joining who are joining us live via the web. I'd like to invite the people who are on the web to email questions to us and we'd be happy to discuss them.

With that, I'd like to introduce the first speaker, Maitreyee Hazarika, who's going to talk about regulatory considerations for alternative endpoints.

Thank you.

[00:02:30]

Hazarika: Thank you, Suzanne.

Good morning, everyone. My name is Maitreyee Hazarika. I'm a medical reviewer
[00:03:00] in the melanoma and sarcoma team. I will provide an overview of the regulatory pathways for approval and the considerations for alternate endpoints.

I have no disclosures.

FDA marketing approval requires substantial evidence of effectiveness in adequate and well-controlled studies with acceptable safety. FDA examines the evidence in the context of the disease, line of therapy, available therapy, study design,

[00:03:30] endpoints, and magnitude of evidence. The aim is the ability to generate product labeling that defines appropriate patient population for whom the drug is indicated and provides adequate information to enable safe and effective use.

There are two pathways for approval. Regular approval is based on the demonstration of direct clinical benefit or effect on an established surrogate.
[00:04:00] Accelerated approval has the intent to treat serious or life-threatening illnesses. Approval is based on a surrogate or intermediate endpoint reasonably likely to predict clinical benefit. Unlike regular approval where there is no requirement of comparative efficacy, accelerated approval provides meaningful therapeutic benefit over available therapy and requires complimentary clinical studies to verify
[00:04:30] and/or describe the actual, clinical benefit.

Accelerated approval is an expedited program. The goals of the FDA expedited programs is to address an unmet medical need in the treatment of a serious or life-threatening condition, facilitate and expedite development and review of promising new drugs, and show that therapies for serious conditions are approved
[00:05:00] and available to patients as soon as it can be concluded that the therapies benefits justify the risks and allows for all the attention to investigation agents that have promised in treating such conditions, including earlier consultation with FDA.

These are the other expedited programs. All the expedited programs are intended to treat a serious condition. Fast track can be based on non-clinical or clinical data
[00:05:30] which demonstrates the potential to address and unmet need. Breakthrough therapy, which is a relatively new designation, is based on preliminary clinical evidence that demonstrates substantial improvement over available therapy on one or more clinically significant endpoints. It is eligible for equal and frequent interactions with the FDA for intensive guidance, including senior management
[00:06:00] guidance. Priority review is granted if the drug would provide a significant improvement in safety and effectiveness and results in a shorter clock end-time to approval. Although a special protocol assessment shown here is not an expedited program, it is another regulatory pathway that allows agreement on clinical protocols, including the endpoints for phase three studies forming the primary
[00:06:30] basis for demonstrating efficacy, foreign [inaudible 00:06:32].

The core development of two or more investigational drugs follow the same regulatory pathways that have been described above but it has additional issues that need to be considered, including combinations intended to treat a serious disease or condition, compelling evidence, biological rationale for use of the
[00:07:00] combination, the contribution of individual agents to the treatment effect of the agents used in the combination, safety evaluation for potential increased risks, and early and frequent interaction with FDA to facilitate development.

Endpoints which have supported past approvals, the so-called traditional endpoints, include overall survival, which is a direct evidence of clinical benefit.
[00:07:30] They require randomized control studies. They can be superiority or non-inferiority survival studies. They tend to have a larger samples size with longer follow-up and

may be confounded by crossover and subsequent therapy.

[00:08:00] Tumor assessment endpoints are surrogate endpoints for clinical benefit, which support accelerated approval. They can also support regular approval depending on the magnitude of the effect, effect duration, disease setting, and risk benefit to available therapy. The endpoints of disease-free survival and relapse-free survival have supported regular approval in the adjuvant setting after definitive surgery or radiotherapy. While objective response rates can be assessed in a single study, the

[00:08:30] other endpoints require randomized, controlled studies. These endpoints have a smaller sample size and some may require blind and external review or audit and the clinical significance is based on the benefit versus risk. Please note that the response and progression endpoints for regulatory purposes are based on conventional or standardized criteria. For example, the recist criteria used for solid

[00:09:00] tumors.

Patient-reported outcomes endpoints are based on quality of life, physical functioning, tumor-related symptoms, and symptomatic-adverse events. These are patient perspectives of direct clinical benefit. They require randomized control studies which need to be blinded and they need a fit-for-purpose tool for the

[00:09:30] measurement of the patient reported outcomes. Missing or incomplete date must be at a minimum and the interpretation of clinical significance of small changes can be challenging.

FDA has regulatory flexibility as states in the regulations and FDA applies flexibility regarding the evidence required to support approval to address particular

[00:10:00] challenges posed by each disease. Over the years, innovative endpoints have supported approval

Section 1 of 45 [00:00:00 - 00:10:04]

Section 2 of 45 [00:10:00 - 00:20:04] *(NOTE: speaker names may be different in each section)*

Hazarika: Innovative endpoints have supported approval. Examples are pathologic complete response which supported accelerated approval for pertuzumab in neoadjuvant breast cancer. Spleen volume reduction was a primary endpoint based on radiographic response rate and patient reported outcomes defined in the total

[00:10:30] symptom score was a secondary endpoint and these two together supported regular approval for ruxolitinib in myelofibrosis. Red blood cell transfusion independence supported accelerated approval for lenalidomide in low and intermediate risk myelodysplastic syndrome with the 5q deletion cytogenetic abnormalities. Major cytogenetic response and major hematologic response

[00:11:00] supported accelerated approval for imatinib and nilotinib in chronic myeloid leukemia.

The need to consider alternative endpoints for immuno-oncology drugs began with melanoma. Ipilimumab the anti-CTLA-4 antibody was approved in 2011. Objective response rates were very low, 11% in a previously treated advanced melanoma

[00:11:30] population. However, it did translate to a benefit in overall survival with reduction in deaths of 34%. Similar findings were observed in previously untreated advanced

melanoma population where ipilimumab was compared to dacarbazine. The Kaplan-Meier curves for overall survival show a delayed separation of the curves and a late plateau with the tail at the end. In this case tumor response based
[00:12:00] endpoints which can be measured earlier did not adequately capture the anti-tumor activity of this drug and would not have led to a demonstration of effectiveness without the results of the overall survival.

With targeted inhibitors of the oncogenic driver mutation BRAF and downstream effectors MEK inhibitors which were approved soon after, there is clinical antitumor activity early on with higher magnitude of responses then approved to
[00:12:30] form 48%-77%. Subsequently immune checkpoint inhibitors the anti-PD-1 antibodies pembrolizumab and nivolumab were approved. Kaplan-Meier curves for progression free survival again showed delayed separation and tail in the curves similar to ipilimumab. The initial approvals of these two agents were based on
[00:13:00] objective responses of 24-32% which are low but with response durations of more than 10 months and even in a phase 1 study with longer follow ups response durations of 23 months were observed. These results were in a very refractory population because all of these patients had been treated with ipilimumab and those who were BRAF inhibitor and with BRAF inhibitor if they were BRAF V600 mutation positive.

[00:13:30] The considerations with immuno-oncology drug development are low objective response rates with prolonged duration of responses which may reflect the dynamics of the immune system. The magnitude of median PFS as shown previously may not represent the treatment effect if benefits are delayed. Median overall survival may represent long term benefits in a minority of patients. Delayed
[00:14:00] separation of the progression free survival and overall survival Kaplan-Meyer curves with late plateau of survival curves which probably represent long term responders who remain progression free for years. The characteristics of the long term responders have not been identified regarding PD-L1 expression although it has been around for some years there remains uncertainty of the value of PD-L1 expression as a predictive biomarker across various tumors.

[00:14:30] The challenges we face include identification of measures of direct clinical benefit with innovative endpoints and identification of surrogate endpoints for overall survival to predict long term outcomes. Examples would be in intermediate endpoints, you will hear some of this in the talks today. To characterize clinical benefit with alternate analysis although the immune related response criteria was
[00:15:00] initially evaluated in patients with advanced melanoma with ipilimumab the criteria has not been sufficiently evaluated for immunotherapies with other mechanisms of action or for other tumor types. Hence it is currently considered an exploratory endpoint for regulatory purposes.

Improvement on the tail of the survival curves can be done by using antitumor
[00:15:30] strategies by use of combination approaches of different immuno-oncology drugs or pairing immuno-oncology drugs with chemotherapy or radiotherapy or sequencing of therapies and using predictive biomarkers all with keeping an eye on

[00:16:00] the toxicity where combinations are concerned. Although the complexity pf immunotherapy responses have made biomarker evaluation challenging. Finally to accelerate immuno-oncology drug development study designs to expedite approval such as enrichment strategies which may be biomarker enriched or biomarker tailored studies in small study populations or adaptive trials.

Thank you to my colleagues for their input and thank you for your attention.

[00:16:30]

Suzanne Topalian: Thank you and now I'd like to introduce Rajeshwari Sridhara who's going to continue the discussion from the FDA viewpoint about challenges in regulating trials based on traditional and nontraditional endpoints in I-O.

[00:17:00]

Sridhara: Good morning everybody. My feeling is today will be more interesting than yesterday at least that's my bias but we'll see how it will go.

[00:17:30] I'm Raji Sridhara, I'm the Division Director of Division of Biometrics in FDA and my division covers all of the oncology/hematology products. We have close to 40 statisticians in my group who cover all of these areas. You'll hear some of the talks in the afternoon from my group as well.

[00:18:00] I don't have answers so I'm only going to present the challenges that we have in looking at some of these data. I have no disclosures. As we all know the traditional endpoints so this is our base. Where are we starting? With objective response rate based on RECIST criteria, progression free survival defined as time from randomization to either disease progression or death, whichever occurs first. Again based on RECIST criteria most often. Sometimes clinical progression is also included in this but however very rarely and it's all based on RECIST criteria. Overall survival itself is of course considered the golden endpoint or the standard endpoint and it's

[00:18:30] timed from randomization to death. Patient reported outcomes have not been used very often but certainly there is a regulatory pathway to using patient reported outcomes where you can show improvement in disease symptoms or deterioration in some of the symptoms as well.

[00:19:00] The disease pathway process in the cytotoxic paradigm has been usually to use tumor response is that it's directly in the pathway and then that results in an improvement in overall survival. However not many studies have shown actually that a difference in response rate leads to a difference in survival. There could be other pathways where you can see beneficial or harmful effects which are not measured directly by the tumor itself. If nothing else, this certainly tells us about

[00:19:30] the activity of the drug and it's measured quite well. However in the immuno-oncology products what we have is the immune response in between which is triggering the tumor response and it could be that even the immunotherapy directly also works on the tumor itself and if we have other biomarkers then they

[00:20:00] could be there in the disease pathway as well. It gets complicated if we are just measuring the objective.

Section 2 of 45 [00:10:00 - 00:20:04]

Section 3 of 45 [00:20:00 - 00:30:04] *(NOTE: speaker names may be different in each section)*

Sridhara: Gets complicated if we are just measuring the objective tumor response rate based on just the tumor measurements. Some of this Dr. Hazarika has already gone through, so our regulatory considerations are based on adequate and well controlled studies, assuming that when we say investigations that we have a

[00:20:30] replication of the effect, and also all this is to see if the drug will have the effect that's claimed by its labeling. In this regard, as statisticians, we have interpreted this law that it has to be statistically significant, most often the five percent criteria is used, but in the law there is nowhere written that it has to be five percent.

[00:21:00] In Oncology studies, as you're well aware, only single studies are used often. The very early Phase one-two studies are sometimes used as the second replicated study. Within the single study we often look at consistency among end points, and this has been a challenge in the immunotherapy setting. The trial designs that are often used are very broadly just single armed or you would have a randomized

[00:21:30] clinical trial. In the single arm trial what we have seen is where the approvals are given are based on patient populations which are more refractory, but we have to be aware that it's difficult to attribute safety concerns when you have a single arm study. Whether it is because of the disease or the drug it's difficult to differentiate in a single arm study. Long term safety's unknown survival are generally short and

[00:22:00] approval that way we have often used accelerated apprual, however in some hematologic malignancies are even in solid tumors where it is a very refractory setting and if you have complete responses this has resulted in full apprual as well.

So in the randomized clinical trials do we have optimal dose? Often we don't. You

[00:22:30] still see in the confirmatory, the so-called confirmatory studies that there are a lot of those modifications as well as discontinuation et cetera. What is the optimum length of follow up? We never pay attention to this most of the time and we go by the even driven analysis, and whenever the events occur we look at it and we don't

[00:23:00] know if we have adequate follow up to assess the safety. We are more and more seeing with these breakthrough therapies that the products are being approved on interim analysis, and so again the follow up in that case is not enough, and we may not have a robust estimate of the evidence. While there is an effect, we may not have a good estimate of the effect.

The optimal length of the treatment is often a question. Treating disease,

[00:23:30] progression, do they need a break in between, would that be better? Those questions are not answered. Approval[inaudible 00:23:34] can be both accelerated or regular using randomized control studies. These are the products that have been approved recently in the immunotherapy space. What I want to show here, in the

[00:24:00] first study when we were just looking at EP the response rate was 5.7%, but then in subsequent studies the same EP gave a little over 10%. There is some variability and perhaps due to patient population or what have you, but you do see this. Among pembro and nevo studies you see that monotherapy is from 20 to 35

[00:24:30] percent around that, but the response rate with the combination is 60 percent. Of

course you have the PFS values which go all the way and we have seen in many studies that there is a survival advantage. Can we drop using the resist criteria when you are seeing such high response rates? We don't know that, so the combinations are still to be determined whether we should continue with the resist
[00:25:00] criteria or if we should modify this, or how should we modify this.

If you look at the non-[inaudible 00:25:05] cancer again here too you have 19 percent in the overall population. Generally they were much lower than what we saw in melanoma, but in the pembro in the high PDL1 expressing group you had a higher response rate, so these are not apple to apple comparisons, but in general
[00:25:30] they were lower than what we observed in melanoma. The response rate varies with the disease.

If we go to renal cell now, it is back to where it was with the non-[inaudible 00:25:41] cancer, but the classical Hodgkin's you have a much higher response rate. When you go to the head and neck again it's not that high, and in the urotilliar carcinoma with that again you are seeing a 15 percent response rate. Although
[00:26:00] there is variability in response rate between diseases are we sure that we cannot use Resist anymore, or what else can we use is the big question, and you'll hear a lot about it in the after note. In looking at all of this we see that with the EP there was a clear improvement in overall survival, overall a response rate, a marginal improvement. PFS in all of these studies we see what is called non-proportionality.
[00:26:30] In all of our analysis of time to event we assumed the proportional hazard rates are proportionate to each other in the two arms, and then we can use these analysis. You will hear a lot more about it in the subsequence sessions. Dr. Latimer, as well as Dr. Chen will be going into the actual definition of this non-proportionality and how maybe we can use some of the endpoints.

[00:27:00] PRO disease has not been evaluated and no long-term experience in at least the IM products. I'm already getting a yellow line there so I better hurry up now. So we have done some work to look at formally whether there is a surrogacy of ORR or PFS among these. The caveat is that because we had limited number of studies we
[00:27:30] have included all diseases, all the randomized controlled trials. At the patient level we have tried to look at responder with non-responders with the respect of PFS and OS and looked at the hazard ratio of that using Copland-Meyers estimates et cetera. We have looked at Spiramen-Rank correlation, it's a simple correlation between PFS and OS based on individuals. At the trial level we have looked at the Arts Ratio based on objective response rate with the overall survival and then Arts
[00:28:00] Ratio with the ratio of PFS. This was done using a weighted linear regression model, the weights were equal to the number of patients in each of these studies, and we included all of the studies. You see that there are more studies here than the trials we have done. Some of the studies had three arms and we have considered them
[00:28:30] as two studies so we can get them.

We had reasonable number of sample sizes in each of these and the overall survival, as you can see, varies from about point 4 to point 9. PFS as well goes from about point 4 to about point 9, and the Arts ratio also varies from quite a bit. At the

[00:29:00] trial-level association, there is no association. I can slow down, I have the green light back again that's good. Here each of these is a different study and the size of the circle is dependent on the size of the study. Again the caveat is that we have put all of the diseases together and maybe that's the reason we are getting such a small off-square radio. Off-square equal to one means a perfect association and
[00:29:30] you can see where we are on that scale. We are closer to zero than to one. We tried using a locked trans[inaudible 00:29:38] hazard ratio as well and still you don't get that. If we look at the response rate of Arts Ratio where it says "the overall survival" there is an outlier here and even if we remove that, this doesn't change much. Again the R squared is very low, so it's telling us really that response rate
[00:30:00] and PFS are not surrogate in points.

Section 3 of 45 [00:20:00 - 00:30:04]

Section 4 of 45 [00:30:00 - 00:40:04] *(NOTE: speaker names may be different in each section)*

Sridhara: PFS are not surrogate end points. Nevertheless, it doesn't mean that we cannot use accelerated approval process. Accelerated approval does not require that it has to be a validated surrogate end point. It has to be an end point that's reasonably likely to predict overall survival. This shows that probably we cannot say that reasonably
[00:30:30] likely we can predict survival based on either of these two end points.

Then we looked at the individual level analysis. If you look at the non-responders, the control and the treatment are smack on top of each other. There is absolutely no difference. But when you look at the difference between responders in the
[00:31:00] treated group versus the responders in the control group, there is a difference. All that means to say that probably the responses that we are seeing, there was 32% to 12.2%. About double the response rate that we got but that is actually adding to the survival. We have already observed that the [inaudible 00:31:22] of response in these IO products are generally longer than what we have seen before.

[00:31:30] This is with respect to overall survival. The same thing was observed with PFS as well. The PFS is longer. I don't know, maybe this was a repeat of that. It looks like I have two same. This should have been a PFS one. No, I do have the PFS, got it. The
[00:32:00] PFS one, again, it is saying that there is really a lot of difference between the responders and the non-responders. There is no question that if we have a higher number of responders in the treated group, then it is going to benefit. We can't still throw away the response idea.

Now, when we look at nivolumab second line [inaudible 00:32:25] lung cancer, you
[00:32:30] have the PFS curve which is crossing. Initially, the treatment was not doing that great and later on. This leads to the argument of whether the progression is being called too early in the IO products and you have some kind of pseudo progression. We are then questioning whether the way PFS is defined is accurate or not. You will
[00:33:00] hear from our group, we have done modifications of this PFS analysis and Dr [inaudible 00:33:03] will be presenting some of this.

With respect to response also, we have looked at this resist criteria response still being there, how can we add the duration of response etc. and that intermediate

end point will be presented by Dr [inaudible 00:33:22] later on today. Here are
[00:33:30] examples of non-proportionality. I think you have seen this graph multiple times already. Essentially, we are saying that all are not separating at the beginning and they don't look somewhat parallel to each other. It is very clear that there is no proportionality. Yet, because we have not come to some kind of consensus on how to summarize these, in all of these there is no question that there is an effect.
[00:34:00] There is an effect, we see that. How can we summarize this? How can we put this in the label is a big question.

Dr [inaudible 00:34:12] went through some of this as well that we have, right now, by a marker of [inaudible 00:34:19] expression is being looked at. Some studies have been done with only high expressers, some with somewhat moderate
[00:34:30] expressers, some with no restriction. In all of them, you see that there is some variability in the response rate. There is a variability in survival but there is no standard SAR threshold that is being used. Are there other biomarkers to be tested? We don't know that. The unanswered question is, we were talking about
[00:35:00] the tail in the previous one where you see all this tail. Is this tail only because there is longer duration of response and that's why you see this or is this a subgroup of patients who really are going to do well?

The point is that you see that in the control arm as well, there is some amount of plateau that happens later on. These could be a separate subgroup of patients who live longer or who have better prognosis or who get more benefit from the
[00:35:30] treatment. We don't know any of these. I'm going backward. We need to know what it is. Can we define a subgroup?

Going forward, what should be the primary end point is the main question that we have. Certainly, overall survival is a clear winner but will it be feasible as we move
[00:36:00] forward? In fast line treatments, if we have more and more effective drugs, which is what we want, in a reasonable period of time, can we prove that a drug has a survival improvement? We are being practical, thinking ahead of what may come up.

PFS is the resist progression criteria accurate for immuno-oncology products or
[00:36:30] should we be using some others? In most cases we have seen this. I have to say that this is not a unique phenomenon. We have seen this pattern even in some cytotoxic as well as TKIs. Somewhat, we have ignored it and said overall, there is an effect. Still, we are putting the hazard ratio in the product label which doesn't make sense.

[00:37:00] We have seen that this [inaudible 00:36:57] response rate is inconsistent. Meaning that some are low responses, some are high responses. We cannot say that for all IO products we are going to use response rate. Should we do it by disease then? Don't know. What happens when we combine these also? The length of follow-up, as I mentioned about it, these toxicities are different from chemotherapy and
[00:37:30] therefore, how long should be the follow-up. We are, of course, evaluating both efficacy and safety. If we are using interim analysis data then how do we get that

safety data? Beyond treatment follow-up, usually the treatment follow-up stops so after the treatment is finished, how much of follow-up?

[00:38:00] Future trial designs in combination with another IO product or with other targeted or cytotoxic therapies, we can't just simply abandon them, the resist criteria that has been used. I don't say that resist criteria has the best anyway for response rate. Maybe it's not the best but that's the response criteria we have used and we have something to compare of the previous products that have been approved.

[00:38:30] Again, the follow-up. The cytotoxic therapies are shorter. How do we then manage this? An end point, toxicity, follow-up, length of treatment, all these are important in the design consideration itself and how we are going to assess these endpoints. One thing that I did not mention, to complicate all this, we have the switch over of treatments. Meaning that the control [inaudible 00:38:53] may get the treatment
[00:39:00] after they progress. In some of the protocols, it is written that they can get the experimental treatment after they have disease progression. Then again, the tail is complicated. Is the non-proportionality of the tail because of the treatment later on or what it is? The subsequent therapies do matter.

[00:39:30] In summary, the current knowledge is limited and we need to explore intermediate end points other than overall response rate and PFS. You will hear some of this in the afternoon. We do need standardization in biomarker measurement and assess. Identification of potential subgroup with long term benefit, I think that's something
[00:40:00] that we all have to try and see. So far, of course, we have tried to look at them and we have not been able to ...

Section 4 of 45 [00:30:00 - 00:40:04]

Section 5 of 45 [00:40:00 - 00:50:04] *(NOTE: speaker names may be different in each section)*

Sridhara: ... We have tried to look at them and we have not been able to identify, but that doesn't mean that in the future we'll not be able to identify that, and alternate ways in which data can be summarized to address this non-proportionality, and again, some parts on this will be presented in the afternoon sessions.

Critical consideration on duration of treatment and length of follow-up is
[00:40:30] necessary. All this relate to some kind of innovative trial designs where we can account for all of this. Our pitch from the FDA is when you have so many products, why not do master protocols and have a common control so that patient resources are not infinite, and therefore, we should be careful?

[00:41:00] Enrichment strategies, adaptive designs you will hear again. Some of this in the afternoon sessions could be considered. If network infrastructure is available and resources are there, certainly, these kind of platform trials are very useful.

With that, I'm very much in time. I have six more minutes. Thank you.

[00:41:30]

Suzanne Topalian: Thank you, Raji, for that really outstanding summary of all of the main issues that

we're going to be discussing today. Our hope, now with this convening of all the stakeholders in IO drug development, is that we will come up with a real path forward after the discussions that we're going to have today. With that, I'm going to turn over the podium to Renzo Canetta, who's going to moderate the morning
[00:42:00] session about endpoints.

Renzo Canetta: Thank you and good morning. I just wanted to remind you that this panel will deal with traditional endpoints. Traditional points are objective response rate, disease-
[00:42:30] free or relapse-free, or progression-free survival, and overall survival. Now, since I am an older person, they asked me to give a historical perspective and that's what I'm doing very briefly. This is my disclosure.

I put in this slide, if you accept that the regulatory history of immunology began with the interferon, that we are now celebrating 30 years of immunology and
[00:43:00] cancer. I roughly classified the approved agents, and these are US approved agents, in three categories; cytokine-focused, vaccine-focused, and then the new [journey 00:43:12] of checkpoint inhibitors. It's impressive to speak about the last century and this century.

[00:43:30] I put on this slide all the currently approved indications, whether they are primary or subsequent, for immuno-oncology and immunological agents and I divided the slide. In the upper part, you see the older agents. Now, these are the compounds that have been approved initially for their first approval or subsequent approval for a new indication based on durable objective response rate. Here, objective
[00:44:00] response rate, remember, we're talking about traditional criteria. It's either WHO criteria with a 50% shrinkage of the tumor lesion or the RECIST criteria with a 30% percent shrinkage. In the old era, high response rate, and I should also add the high response rate in the absence of decent alternatives as was the case for the HIV-
[00:44:30] positive Kaposi sarcoma population, we are able to produce approval.

Going on, the concept of durability kicked in and here, as you can see, in the case of BCG, for the in situ tumor, high response rate and durability, they both were remarkable. For interleukin-2, it's interesting that the response rate were not impressive and that also a data pointed to that also for the novel agents, but the
[00:45:00] duration was quite impressive, and I have to say that in the case of a melanoma, the complete response duration achieved with interleukin-2 is the one that really won the day.

As we moved on to this century and the recent years, as you can see, the attitude and the possibility that we have in this country to achieve accelerated approval
[00:45:30] based upon response to be confirmed with later results from the same trial or from subsequent trials with other endpoints has been playing a major roll. It's interesting to see that at least in the case of the PD-1 inhibitor, now, the overall survival data start kicking in and the accelerated approval can be converted and have been converted into full approvals.

Now, when you look at progression-free or disease-free survival in the adjuvant

[00:46:00] setting, this is what is being experienced in the past. Again, in the older years, hazard ratio were not used. I have to say that everything that I'm showing is present in their respective labels of the drug with only one exception I'll point out later, but again, here, there were advantages in a number of situation. It's
[00:46:30] interesting to note that for adjuvant melanoma both in the case of the pivotal trial for interferon, and as we [inaudible 00:46:36] the last week from ipilimumab, now we have survival advantage results, but these were not contained in the original submission.

Again, this is what is produced approvals in the old era. In the novel era, disease-
[00:47:00] free survival and progression-free survival have been achieved. Here, I would like you pay attention to the hazard ratios that, again, they're quite impressive even for if you want to define it as a weak endpoint or a non-definitive endpoint, the hazard ratios are quite impressive.

Now, thanks, God, we live in an era where agents can actually prolong survival and we'll see more and more evidence of this prolongation of survival. In the past era,
[00:47:30] only the interferon alfa-2a achieved the survival advantage over chemotherapy in chronic myeloid leukemia. In the more recent years, there have been quite a number of trials that have shown a survival advantage.

Now, just for the fun of it, I would remind you of a paper that ASCO, which is a society I'm a member of, published recently indicating that every drug that
[00:48:00] achieves a hazard below 0.8 for survival should be made available to the public. Keep that in mind. As you can see in all of this cases, that hazard ratio was achieved.

I put on this slide looking at the pivotal trial that achieved the survival advantage asking the question, was there a consistency? You heard Dr. [inaudible 00:48:20] allude into consistency. It's interesting that despite all the limitation of the
[00:48:30] behavior of this agents in a response rate, response rate were actually significantly superior in the experimental arm quite constantly with the only exception of chronic myeloid leukemia where hydroxyurea and busulfan did actually do better than interferon for objective response.

Now, one can argue the perhaps because of pseudo-progression, objective
[00:49:00] response can be underestimated, but I would think twice about saying that they may or may not be reliable, so something to keep in mind. However, for progression-free survival, there have been inconsistencies and one of them was pointed out by the Dr. [inaudible 00:49:17]. I'm pointing out also to this one. Actually, in this particular case, more than one trial had the lack of difference in
[00:49:30] progression-free survival between the experimental agent and the control.

Now, keeping this in mind, these are the questions that we will ask during the panel and I would invite the people that are attending by computer also to send in their question. We have a panel of experts that will deal with traditional endpoint in terms of imaging, in terms of statistical challenges, and in terms of clinical

[00:50:00] responses. We start with imaging, we start with Dr. Schwartz from Columbia University.

Section 5 of 45 [00:40:00 - 00:50:04]

Section 6 of 45 [00:50:00 - 01:00:04] *(NOTE: speaker names may be different in each section)*

Renzo Canetta: ... From Columbia University.

[00:50:30]

Schwartz: Thanks. It's a pleasure to be here, and I appreciate the invitation from the organizers. I will focus on the imaging aspects of response assessment in immunotherapy. Really, what's brought this to everybody's attention is, early on, in treating patients with these agents, we saw, sometimes, an atypical pattern of response, whereby initial growth of target lesions, or the development of transient new lesions, followed by improvement, so what would have traditionally been

[00:51:00] called progression, we recognize by the mechanisms of active immunotherapy, may actually not represent progression, at all. Additionally, the concept of new lesions really changed for us. Small lesions usually, quite frankly, below the resolution of CT scan, may appear as new lesions, usually on the first or second follow-up, which do not progress, and they ultimately, in fact, decrease. This actually, again, represents a response, instead of our usual definition of progression.

[00:51:30] RECIST 1.1 had language that describes how to manage equivocal new lesions, but this was not based on immunotherapy, and in fact, may not actually be appropriate for immunotherapy. Several different approaches have been proposed. The so-called immune-related response criteria, which is really more of a guideline than a criteria, had been very, very helpful in the beginning, and we're seeing, and one of the things that I'll talk about now, is how we have to come together on this and

[00:52:00] really have more than just a guideline, but actually a criteria of very specific rules that we can apply to the patients on these trials.

I'll show some examples of these, since we are talking so much about response and progression, and actually what they look like. I'll start with a lymph node. You see this in the right groin at baseline. Cycle 2, it grows. It would be considered progressive disease, but in fact, by cycle 4, it's regressed again, so a classic example

[00:52:30] of a flare. I start with a lymph node, because that was the organ that it was expected to occur in, or people thought that it would just be in lymph nodes, but in fact, we see this really in all sites of metastatic disease.

Here's an example of a lung lesion. Here, by cycle 2, it grew, and then regressed, and actually represents a response. Notice this case, and there's going to be a second finding in each of these cases. Notice this case, there's a second lesion in

[00:53:00] the right lung, which actually didn't have the flare, and actually immediately responded.

Similarly, this para-renal mass, where again, we see growth at cycle 2, regression at cycle 6, but it would not be a response at cycle 6. It still remains larger in size than at baseline, but in fact, it's less dense. It's more necrotic, so is that a different type

of response that we have to capture?

[00:53:30] A very interesting case, these new subcutaneous nodules at cycle 2, which regressed, thought to represent a flare, and then progressed, or appeared again. The patient was called progression at cycle 8. I'm wondering, could this be a secondary flare phenomena that we're seeing in these same lesions, or is this true progression?

[00:54:00] Finally, a hepatic metastases. What I want to draw your attention to here is certainly the excellent response by cycle 6, after an initial flare, but areas of enhancement within the lesion, for which the referring oncologist, or the treating oncologist, truly felt that this may represent progressive disease.

[00:54:30] Then, the other challenge is the time of progression, or the time of pseudo-progression. Here's a classic example of a large groin lesion at cycle 2, but then rapidly shrinks by cycle 4 and 6. Here's another interesting case of a subcutaneous lesion in the right back here at baseline, actually responds at cycle 2, flares at cycle 4, and then regresses or responds by cycle 6.

[00:55:00] These are true challenges that we have to address. I would say, though, for each of these cases that I show of pseudo-progression, we see many, many cases of true progression, and somehow, we have to distinguish the 2. I say many, many cases, because in fact, that's one of our first challenges. We actually don't know how frequently this phenomena occurs, either based upon the indication, the primary tumor type, or based upon the specific immunotherapy. There are truly differences with different indications, and with different immunotherapeutic agents.

[00:55:30] I do want everybody to recall that, and I'll speak a bit more now about response and progression as distinctive events, in terms of treating patients, and in terms of clinical trials. Response and progression really measure 2 different things. Response early in the course of treatment doesn't really have a true role in clinical practice, and this is why we speak, and we've heard already, so much about post-progression scans and collecting them, and rolling clinical research, obviously, primarily to calculate an overall response rate, whereas progression is assessed at intervals until the end of therapy, commonly used to determine when to change therapy, so the need to collect post-progression scans is one of the points that I'll try to make frequently in the next 9 minutes.

[00:56:30] What are the challenges? We've already heard some of these. As market access to immunotherapies increase, overall survival will be increasingly confounded as a primary end point, so PFS will become increasingly important, or some metric related to PFS or response may be. In fact, recent analyses of randomized trials have indicated that there may be an improvement in OS with minimal or no improvement with PFS as assessed, and we heard already by RECIST 1.1, but irRECIST like, or modifications to RECIST, to, quite frequently, take into account some component of this flare may actually correlate better.

[00:57:00] What's important to note is that the original IR criteria, now adopted in several different ways, and that's one of our other problems, that it's been adopted differently by different sponsors, truly was created as a straw man, as an attempt to solve this limited problem, an attempt to address this limited problem, but we actually have to figure out a better way to solve it. What we do need is a uniform definition of irRECIST, which is based, actually, on data.

[00:57:30] Another important point that I'd like to make is that this can only be accomplished by analyzing the actual images. What we've talked about a lot are the case report form-based analyses of response and progression, and we'll hear a lot more about that, I believe, in the next session, and even in the afternoon. This is very, very important, and certainly gives us a potential indication, or an inkling, of how different metrics would perform, but especially for immunotherapy, and especially [00:58:00] for the challenges that I've described, imaging post-progression and things like that. These case report forms are filled out for one purpose, and one purpose only, for assessing RECIST response. It's very good at doing that, but in fact, the data that's collected, and the limited data that's collected, is actually not designed to capture progression phenotypes or kinetics. We don't even capture true tumor burden. We heard a little bit about that yesterday. We're capturing a tiny subset of the actual [00:58:30] lesions, and how that correlates with the total tumor burden that the patient has is really not known, and limited studies have been done, show that there's relatively little correlation.

[00:59:00] Since progression is so important, I'd like to point out this limitation, incomplete non-target and new lesion evaluation. I think it's important that we just say this and recognize this. Being one of the Reference Radiologists at Columbia who frequently fill out these forms, once we generally see 1 reason for progression, unless we're specifically asked to complete the form differently, we may not record the other lesions' reason for progression. If we have target lesion progression, we may not look at all the other sites of disease. Now if you, and I believe we should ask for all those sites of disease, as I'll show you, because I think this could help our PFS OS correlations quite a bit. Post-progression measurements are inadequate and, quite [00:59:30] frankly, inaccurate, because they're not included on this form, and I've seen many, many valiant attempts to get them from radiology, actual radiology reports, that are read by somebody else. I think all those valiant attempts are wonderful, but quite frankly, they're usually inaccurate.

[01:00:00] Are we adequately accessing response and progression? We, and others, including regulatory authorities, have shown ... This was a meta-analysis that we published in JAMA Oncology a few months ago, that increase-

Section 6 of 45 [00:50:00 - 01:00:04]

Section 7 of 45 [01:00:00 - 01:10:04] *(NOTE: speaker names may be different in each section)*

Schwartz: ...analysis, that we've probably seen in JAMA and Oncology a few months ago that increased overall response rate, was found to be significantly associated with a likelihood of regulatory approval and clinical benefit. It was strongest for single agent regimens but overall response rate is just as we heard, one of many metrics

that we could look at and there probably are better response metrics given that we
[01:00:30] have the data, certainly the images that we could look at and I've just listed some examples here.

Similarly, are we adequately assessing progression? And I'd say this is an area that we're clearly inadequately assessing. This is a slide that Geoff Oxnard, many of you know, showed at ASCO in 2012. This is a patient with lung cancer on a TKI, started therapy here at base line. By 3 months, had a response in these lesions in the right
[01:01:00] lung and by 14 months was a RECIST PD. But notice that in fact, despite the fact that they were RECIST PD, based on growth from nadir, the tumor size is actually less than what it was at base line and he and his colleagues actually continued this patient for 39 months. So, 25 months greater than 2 years post RECIST progression, without clinical symptomatology. And was only at 39 months that the patient first
[01:01:30] had dyspnea. So does progressive disease as defined by many of our criteria actually correlate with treatment failure? Probably not, as we define it.

And then you know, I think I have it in the next slide, that's fine, are we adequately assessing progression? I'll show you that same patient again here, example 1, and
[01:02:00] another patient with lung cancer that had growth in a target lesion of greater than 20 percent as well as a new brain metastasis.

Both these patients would be considered progressive disease based upon our characteristics and our 3 categories of progression. However, I could assure you that the OS between these 2 patients is significantly different. This patient with the edema, with the enhancing brain metastasis, did significantly worse.

[01:02:30] So we've seen you know, that we have this challenge with progression. We also have variable interpretations of progressive disease with immunotherapy. In target lesions, some of diameters exceeds PD threshold but how do we actually, to take into account the flare phenomenon, how do we actually confirm that. Is it continued growth, is it further growth, is it continued PD alone, is it further growth,
[01:03:00] do we reset to a new base line and have to progress from that point.

Similarly, with non-target lesions, a rather subjective criteria of neuron equivocal progression. Is it clearly unequivocal. Is it only if there's target lesion progression. Does further growth beyond unconfirmed PD indicated. And again, these are all paradigms that different sponsors have adopted in order to standardize it but now
[01:03:30] we a little bit of a Tower of Babel with different standards between different sponsors.

What we really want to do, again certainly to validate this and to ascertain it's accuracy, is to standardize. But what we really want to do optimize it, is to define a progression phenotype. Tumor growth rate from prior scanner from nadir would be a phenotype. Is there a gradation of progression phenotypes associated with
[01:04:00] specific post-progression outcomes? Or do we look at the post-progression scans to study the association of these progression phenotypes, be it a new lesion, be it rapid growth of non-target lesion or a certain different kinetic of the target lesions

with dramatic progression or stabilization on post-progression scans. And I think that's how this is some of the different ways that we can make that PFS OS
[01:04:30] correlations, actually work. Both on the patient level as well as on the trial level, better.

So I show this case of a patient with melanoma and a liver metastasis here which on cycle 2 and this may be a little bit too detailed. But new lesions which were probably present but just flared but I tell you, depending upon how many of these I
[01:05:00] measure, how I measure them, what I considered happened to this lesion, this patient could be considered based on our criteria and based upon the variability in the criteria, stable disease or progressive disease. And clearly there's been regression from this point so the patient should be considered stable disease but I could also confirm progressive disease again based upon our rules.

So these are great guidelines but they are really more based on instinct rather than
[01:05:30] outcomes correlation and being data driven. We need analysis of existing imaging, and I mean the actual images that are going to be re-analyzed for the purposes of the new criteria to standardize and optimize. So it's both standardization and optimization of any type of new criteria.

RECIST provides us with a historical basis. It's simple, it's a great tool. IR RECIST is a
[01:06:00] work-around for pseudo-progression and it certainly is a better reflective of benefit with many of the immunotherapies. But now's the time I think for us to optimize this with novel metrics that have statistical strength, independent of post-progression scans and then we can evaluate and analyze prospectively, potentially reducing trial sizes as well.

So what are some of these alternate analyses? It starts first with a comparison of
[01:06:30] RECIST and IR RECIST. Then it should look at optimization of IR RECIST, comparison of response and progression kinetics per arm of a trial, identification of metrics which can consistently predict OS benefit with a fraction of the patients, development of progression features predicting pseudo-progression and then selection of a leading efficacy metric and we hope to integrate that prospectively into immunotherapies, first as a supplementary endpoint or as a primarily endpoint
[01:07:00] if the regulatory authorities let us.

What else can we do with these images? And again, as directed I limited the talk only to existing endpoints and I stretched that just a little bit to existing modalities that we're obtaining. So radiomics is a relatively new technique in imaging. Quantitative molecular imaging provides a potential platform for linking specific imaging traits with a specific gene expression patterns. And these imaging features
[01:07:30] may actually serve as molecular surrogates that contribute to diagnosis, prognosis and likely gene expression associated treatment response.

So here's a patient on a TKI, and one thing that we noted, that again that these patients didn't have the classic response pattern. Certainly not the depth of response that was assessed but actually the tumors became ghosted and this is a

[01:08:00] way of looking at the individual pixels in here to define a response phenotype based upon certain features in this. Basically the graying of the tumor that was present in patients that are EGFR mutant, who are receiving the TKI, potentially with the exact same scans that we had, we may be able to distinguish between true progression and pseudo-progression based upon the patterns that we actually see in the individual pixels on the same imaging set.

[01:08:30] So in conclusion, what we need is better data collection of the actual images and the appropriate meta data and this should include images that are obtained post-progression. That's absolutely critical. We need to standardize IR RECIST as a first step with very specific rules of response and progression. And then we also have to enhance the data collection so our case report forms are more robust and
[01:09:00] complete. For instance, just beginning with the basic concept of record all the sites of disease and all the different methods that we see progression. And then we should explore alternatives. Radiomics is one that I briefly touched upon solely because it's available. It's actually available on the actual images that we have. I wouldn't be a radiologist without mentioning other imaging modalities, my colleagues would kill me. And that's all I'll say about that. But you know, we have
[01:09:30] PET and MRI and other and certain tracers that can help a lot.

Thank you very much.

Renzo Canetta: Usually we have the question at the end of the panel but if it is fairly burning, you can ask. It's not burning. Next Dr. Sumithra Mandrekar from the Mayo Clinic group.
[01:10:00] She's still focused on the RECIST challenges in the

Section 7 of 45 [01:00:00 - 01:10:04]

Section 8 of 45 [01:10:00 - 01:20:04] *(NOTE: speaker names may be different in each section)*

Renzo Canetta: A focus on the RECIST challenges in assessment of outcome.

Mandrekar: Good morning, and thank you for the opportunity to be here to share some of my
[01:10:30] experiences and lessons that we learned using the RECIST database, but really traditional endpoints and traditional agents. This doesn't have anything to do with immunotherapy, but I guess we can learn some of the lessons that we learned, based on the RECIST experience.

I'd like to first acknowledge all my collaborators that worked very closely with me and the rest of our team on this work, and in particular, I'd like to acknowledge my mentor, Dan Sargent, as many of you know, who passed away a few weeks back.
[01:11:00] He was instrumental in actually a lot of this, a lot of this research, and so I'd like to dedicate this talk to him, and I also see Jan in the audience. Jan was very graceful to share these data with us from the RECIST data warehouse, and was a consultant and collaborator in this work as well.

Some of the challenges and successes that we encountered when we analyze data
[01:11:30] using the RECIST data warehouse is what I'll first show, and then what are the results, and what did we learn? Of course, everybody has already talked about it,

response using RECIST is relatively simple to determine, and so it's easy in terms of clinical trial decision making. It's also easy for patients to understand that their tumor is shrinking, how much has it shrunk, so there's definitely a lot of value to having a metric that is simple.

[01:12:00] I guess the challenges that we encountered as we try to analyze these data and trying to see if we can come up with alternative metrics. Some of these things I'll highlight. I think this is just common, because of the way clinical trials are done. I don't think this is anything to do with a particular trial, but it is just really how clinical trials are done. You will have missing measurements, because the patients are not well, they have to travel a long distance, so they just don't show up for a particular assessment. You have inconsistently measured lesions. What that means is you have a set of lesions that you assess a baseline, but you don't necessarily
[01:12:30] record a measurement on all of those lesions is subsequent assessments.

Sometimes you can have conflicting measurements, not exactly at the same time point, but let's say you are trying to, for some reason, this patient comes in and has multiple assessments within a two week window or something, and then you're trying to see, what was the response at a particular time point? Then you have multiple assessments within the time window. Then, of course, I'll show you that there were a lot of data that was based only on clinical examination, so when
[01:13:00] you're trying to come up with new metrics with measurements, it's very hard, if I have only a clinical examination and is this a response or not, I don't have an actual measurement associated with that. Then the response rate itself, it doesn't capture cytostatic activity, and we all know, we have talked about this many, many times, you're categorizing a continuous value, so there's definitely less information.

[01:13:30] Then, of course, there's the issue of okay, then can we use continuous metrics? Of course, but then you have progression due to new lesions, at which time I don't actually have a measurement, so how do I give a measurement to that time point, where this patient progressed only because of new lesions? You can impute, you can do all sorts of things, but this is a reality that you have to deal with when you deal with these data.

[01:14:00] We had over 8000 patients in this database, from three different tumor types, and you can see, breast, non-small cell lung cancer and colorectal cancer, and right off the bat, we couldn't use a lot of these data because there were measurements from clinical evaluations or there was no measurements, because these were non-measurable disease, and then we also excluded patients because you don't have post-baseline measurements. This is possible because the patient comes in, they quickly progress from a new lesion, they die or whatever, and so you don't have any post-baseline assessment. If I'm trying to use, come up with a metric that talks about measurements at baseline and post-baseline, I don't have a post-baseline.

[01:14:30] Then, of course, there's this issue that I talked about, conflicting measurements, and then not all lesions are measured consistently, so if I go down this list, I end up with less than 50%, in some cases a little bit more, of patients that I can actually

use when I'm trying to develop a new metric.

The assessment modality, this is the first thing I was going to talk about. This is the clinical assessment, so we had over 25% of patients with a breast disease that had
[01:15:00] assessments based only on clinical evaluations, all assessments, or at least one assessment based on clinical evaluation. This becomes problematic when you are trying to develop a new metric based on data. Progression reason, there's a number of reasons they have progressions. The target lesion, target and new lesion, but if I look at just this row where you have over 23% and 12 and 8, 10% of patients that actually have progressed because of a new lesion, which means I don't have a measurement at the time they reported disease progression.

[01:15:30] Timing of assessments, I think this was, Larry talked about it very briefly. Of course everybody likes more frequent assessments, right? It gives you more real time monitoring about the tumor activity itself, but it is not according to clinical care, that's just, this just gives you a lot more missing data, and then how do you actually deal with it at the back end, when you're analyzing these information? The studies that we used, they all had protocol assessment times. You see most of them are
[01:16:00] every six weeks, some of them are every three weeks, and all the way up to 24 weeks. When you just look at the post-baseline assessments per patient, very soon, when you go to the patients with more than 3, 4, or 5 assessments within the 12 weeks that produces rapidly, and if you start looking at 24 weeks, again, we have very few patients that have measurements when you get that long out. When you're trying to develop metrics based on RECIST criteria or tumor shrinkage, and
[01:16:30] saying, I want to pick a time point at 12 weeks, 18 weeks, and 24 weeks, you probably don't have that many patients remaining that actually started your study.

Assessment window, again, this is something we try to understand, how often are patients coming in based on the scheduled assessments? If their observed assessment time was within plus or minus two weeks, then we calculated, recalculated the next expected assessment window, just to make sure that we are capturing as many patients as we can in the scheduled protocol assessment time. If
[01:17:00] the current expected was four weeks and the next expected was eight weeks, so every four weeks, right, but the current observe was at six weeks and our adjusted window would be ten weeks instead of eight weeks. We try to figure out how many of these patients were coming in at the scheduled assessments, protocol scheduled assessments.

There were many reasons for not being within the window. We call that it's not perfect, but again, this is all clinically valid. The question is, how do I account for that statistically? They go off treatment early for reasons other than progression or
[01:17:30] death, they missed an assessment followed by a progression without an assessment, so they have a new lesion, let's say. Out of all these studies, we classified how many of these patients per study came in within a protocol scheduled assessment windows or not, and so you see about 40-60, roughly, about the percent of patients that had a perfect assessment, which means we were able to get measurements every assessment time as determined by the protocol, versus

not.

[01:18:00] After all this, this is some of the data challenges that we deal with any sort of database, any clinical trial database, but what did we learn, just with the RECIST experience? We looked at a number of multiple metrics. I'm sure some of this is familiar to everybody, but either you categorize response versus other, so really I'm only interested in patients that achieve a PR or a CR, or I want to categorize
[01:18:30] progressive disease versus everybody else, or you consider everybody individually. There's many, many articles on this, and the right way to do it and all that stuff. We tried to use these data to see if they would give us any difference in the predictive ability for OS, if I considered each of these different metrics. What is the value of considering them differently, and is one better in predicting outcomes, overall survival outcomes, or not?

[01:19:00] Obviously this ... The first thing is you consider the patients that had a CR, PR, stable disease, or progressive disease. Now, A, B, and C just refers to the different tumor types we looked at within breast, non-small cell and colorectal cancer. Of course, the patients that have a progressive disease did worse, and we used a landmark analyses, subsequent survival post 12 weeks, so that was consistent across all our disease types, and also between, at 12 weeks and also at 24 weeks post-treatment.

[01:19:30] Having said that, okay, so people with progression, patients that progressed prior to 12 weeks, I think that's just, you just know that they just did poorly, so they're going to do poorly subsequently, but the other important thing that we really wanted to understand is, how do these metrics, whether I look at them at 12 weeks, I look at them at 24 weeks, either is it breast, non-small, colorectal cancer, we split all our data into training and test, you know, validation data sets, I'm not
[01:20:00] going into all the mechanics of how we did it, but you can see.

Section 8 of 45 [01:10:00 - 01:20:04]

Section 9 of 45 [01:20:00 - 01:30:04] *(NOTE: speaker names may be different in each section)*

Mandrekar: I'm not going into all the mechanics of how we did it, but you can see C-index is one of the criteria we used to understand the predicative ability of these metrics on overall survival. Even though the point estimates look like some metrics may have better, point-wise prediction, but actually, if you look at the ninety-five percent conference [inaudible 01:20:24] they're all overlapping. Really, all the work on ... Of
[01:20:30] course there's a clinical rational for not combining patients that are PD and SD together, for example. You have all this data that says, "Did they respond, or not," so you ignore the stable disease. The question is this clinical rationale for splitting these response categories, but all I'm saying is, if you did that, and if you tried to understand, if calling the response status up to twelve weeks or twenty-four weeks, splitting them by any categories here, all I'm saying is it doesn't impact my
[01:21:00] prediction on overall survival.

That was one of the main conclusions from our initial work where we just looked at category [inaudible 01:21:11] metrics. Then we continued to say, "Okay, let's look

at some simple continuous metrics, right?" Because, all the categorization, we don't like it, we want to look at change in tumor size. That's another big thing that everybody wants to look at, so we looked at zero, six weeks, and twelve weeks, either absolute change or relative change.

[01:21:30] Here is our metrics was really C-index, and here is the three tumor types, the disease model, and then we have the slope metric and the percent change metric. These are the point estimates, and this is, what you see here, the purple asterisk is from the RECIST categorization, so this just looks at the tumor status at twelve weeks, or twenty-four weeks, but does not account for the continu- It's not the continuous metrics.

[01:22:00] So all I'm saying here is this scale is pretty small. What you look here, looks like it's too low, but it's not that different. All I'm saying is across the tumor types, right? Whether you use the slope or the percent change, whether you use the RECIST categorization, I don't see a big difference in any of these metrics in their ability to predict survival.

[01:22:30] The purple triangle here, however, is the time dependent model, so in other words we are just using progression as a time-dependent covariant to predict overall survival, so of course the C-index and those were the highest. All this discussion about, "I need to use only continuous metrics, or I need to use stable disease PRCR and PDS individual metrics," I don't know that, our data at least, our work from here, doesn't really suggest that any of them have any difference in its ability to predict survival.

[01:23:00] So, summary: We looked at many different metrics. They all had similar predictive performance, similar categorizations, and we looked at prediction errors. I didn't show you this data, but we did a number of simulations to say, if you start with a positive phase three, you go back and simulate data from phase two, and then look at positive and negative predictive value. We didn't see a difference in any of these metrics, so our conclusion was these are no better than RECIST based tumor responses, phase two and points again, these are phase two end points.

[01:23:30] We didn't find a lot of differences in where you landmark them, is it twelve weeks into treatment? Twenty-four weeks into treatment? Of course, there was something with breasts that none of these metrics really performed as well as they did for colon and non-small cell lung cancer. Pseudo-progression, did we look at that? The way we tried to account for this a little bit was we tried to use different cutoffs for PD and PR based on the data we had. So for PD, we looked at increase in ten percent, fifteen percent, and twenty percent from baseline. You have to realize we don't have data beyond that, right? Because there's only information up until disease progression. The same thing we did with PR. We went all the way from ten

[01:24:00] to fifty percent in five percent increments, and you know that a thirty percent increase, and a twenty percent increase, it corresponds to the RECIST categorization.

There was no change in predictability again. The RECIST cut-points determined similar predictability as an alternate cut-point. Here's another massive [inaudible 01:24:16] that looks at, again, three different tumor types, all the different cut-points have to be examined in the study, just for this metric, which is a tridr. We use different cut-points re-computer the metric and looked at its association with
[01:24:30] overall survival. Again, you see that there's not ... The point estimate maybe higher for some versus the other, but this is where the RECIST one lies, and I don't see a difference in all in where you place the cut-point. We did not have data beyond disease progression, so we haven't analyzed that.

I have thirty seconds. I just want to switch topics. This is what we learned from our RECIST data base, and switch topics because this came up in the follow up from our
[01:25:00] call. What happens to long-term survivalists? This is not immunotherapy data guys. Okay? This is just a different trial. I'm not telling you which trial is which disease, because it hasn't been published. But this is the data that we encountered about a month ago. Of course this has crossing survival curves of classic non-proportional hazards, but it has this long tale that everybody here isn't interested in knowing how to analyze. One of the things we did, naively, is just comparing these two
[01:25:30] curves, using And I know Raji's going to say, "Well that's not proportional," but yeah, so there was no difference. It was clear that there's something else that's going on with these data. We did end up using a mixed rate cure model. These data again, are not published, but hopefully soon. What does this do? This assumes that a population of some [inaudible 01:25:49] individuals, those who have experienced the event of interest and those that have this long curve, I'm sure there's going to be a lot of interesting talks today on this topic. I just wanted to show you one quick example.

[01:26:00] Each is modeled separately, you estimate the care rate for the treatment, and the survival rate for the uncared patients. So, a little bit of statistical jargon, just so I feel good about it, so T is the failure time of interest and you have to model the probability of a patient being cured, given a vector covariance it can be completely independent of the vector covariance for the survival probably different for the uncured patients, right? Then you look at this, and you model both of them, in a mixed rate cure model.

[01:26:30] We fit this model to our data, and we were able to estimate, "What is the cure fraction, so what's the long-term survivals, and what's the cure fraction for arm a and arm b?" After we do this, set of difference between A and B, and yes there is a difference, right? Then we fit a predictive curve from the mixed rate cure model to our data, and you can see, actually in this case, because we have the follow up is
[01:27:00] nine to six months. We have very, very long follow-up in this trial. They've been tracked very good. The cure model, you cannot even see these two colors. This color you cannot even see, it's completely masked by the blue and the red.

So, All I'm trying to say here again, is this data, what we are trying to do in immunotherapy, is there's a lot of learning that we can already do with the data that we have from other traditional agents. I think we need to leverage this data as

much as we can, before we try to come up with new metrics and new ways to
[01:27:30] analyze data. So with that, I will stop. I think I'm a little over, but I'd like to thank the RECIST [inaudible 01:27:34] committee for giving me the opportunity to work with their data, and then all my colleagues will help me graciously with the slides because the scramble the last few weeks. All right, thank you.

[01:28:00]
Speaker 2: Thank you. You're not over at all. I'll introduce now Doctor R[inaudible 01:28:01] whose [inaudible 01:28:03] was instrumental in [inaudible 01:28:05] or in the initial papers that describe the immune-related criteria, and how they apply to resist. A**

Axel Hoos: Thank you Renzo. It's a pleasure being here. I'm happy to make a contribution to this discussion. Actually, from somebody's perspective that has been more than a
[01:28:30] decade already on the topic of immunotherapy [inaudible 01:28:33] it's very nice to see how this is coming together, and how much alignment we already have around the issue, just from the conversation that happened this morning. I'm expecting there will be more of that as today progresses. What I'd like to do is, actually make the bridge now between the known, the conventional, endpoints. The known conventional endpoints that we we'll be talking about in the next session, tell you a
[01:29:00] bit about the history that we have had, tell you about the hypotheses we drew from the early data that told us we need new endpoints, look at the challenges for clinical trials, and how we might improve those endpoints, but not answering all the questions here because the next panel will hopefully do that, and then look at what's the ultimate goal that we're trying to achieve.

So starting that, I make the simple statement about cancer immunotherapy might be somewhat repetitive, but I think it's really important to be clear. The mechanism
[01:29:30] of action as [inaudible 01:29:30] what we do is chemotherapy a targeted agents is just not the same. It's indirect, we're targeting the immune system and not targeting the tumor. The immune response that results from that is dynamic. These are living cells that are doing things. They're not just a dying tumor, from a [inaudible 01:29:51] [inaudible 01:29:52] agent. Having said that, you achieve more than one affect when you activate the immune system. Now the consequences are
[01:30:00] there's no direct anti-tumor affect, and that might be delayed affect

Section 9 of 45 [01:20:00 - 01:30:04]

Section 10 of 45 [01:30:00 - 01:40:04] *(NOTE: speaker names may be different in each section)*

Speaker 1: Direct antitumor effect, and there might be delayed effects that actually show you biologic and pharmacodynamic activity that can be measured well before there is clinical activity and that come in all kinds of shapes and forms. The endpoints we're talking about, they're influenced by that.

Having said this, this is a graph that you are all familiar with. It comes from Bob Schreiber and his group around immuno-editing. It basically makes a simple
[01:30:30] biological case for how the immune system interacts with the tumor. That can then translate to endpoints or observations you make in the clinic. What do we see here? You see the three Es of immuno-editing. Either if the immune system

outperforms the tumor basically the immune response is strong, it can actually lead to elimination of the disease. If you look at an equilibrium, the immune system may keep the tumor in check but it is not eliminating it. That's a very common
[01:31:00] phenomenon with immunotherapy. The last item is of course escape. If the immune system loses the battle, then the tumor will progress and ultimately the patient may die.

If you overlay that on this biologic graph, how we actually translate this into clinical parameters, you may call elimination "response," equilibrium "disease stability," and ultimately progression be the "escape." Very simplistic way to look at it, but that's probably helping to understand the biology as we look at clinical observations.

[01:31:30] Now what are those clinical observations? Let's start with the ipilimumab because that's where the story began. We still have a lot of the data that's being discussed, comes from the early ipilimumab observations.

We have seen some CAT scans already, so I'm not going to get into too much depth here. Larry Schwartz made a very good case for what you can observe radiologically. Here you have a case of a conventional response. There is a lesion at baseline that shrinks, and immunotherapy can do that particularly the PD-1
[01:32:00] blockers do that very nicely.

There are other examples though. The delayed response type phenomena, or pseudo-progression, or flare, as we have named them, they can show that the tumor can get bigger before it shrinks. This is an exam that I like because it's actually not a CAT scan. It's a lesion on the back of a patient that gets very inflamed. Here you can see the quality of the lesion not just the size. The quality
[01:32:30] changes from an indurated nodule to something soft and spongy. There's detritus on the surface of the lesion. It suggests there's activity here. Something is happening to the lesion that probably is induced by the immune system. Ultimately the lesion shrinks and disappears. This patient has become a complete responder for quite a long period of time.

If you want to translate that into a radiograph, here is another example where you have a very massive metastatic lesion in the thigh of a patient that gets larger and
[01:33:00] then shrinks. Again another example of a delayed response. These are phenomena that we see and that we of course want to be able to capture systematically.

Here's another example. Not uncommon at all. Stable disease where over prolonged periods of time, the immune system may create that equilibrium with the tumor and keep the tumor in check, but the lesion remains.

Here's the last example. A new lesion appears while there's a response in existing
[01:33:30] lesion. The new lesion is indicated by an arrow here. It's a very small little lesion, but that lesion of course would under conventional criteria be progressive disease. It means treatments [inaudible 01:33:40] and then ultimately losing the

opportunity for the patient to have more benefit.

If you take all of that and you put it into tumor volume graphs over time, that's what these four patterns look like. A conventional response in the left upper corner. A slowly declining tumor volume, so stable disease in a sense over long
[01:34:00] periods of time, in the right upper corner. A delayed response in the left lower corner. Tumor volume increases before it shrinks. Then, the appearance of a new lesion added to the total tumor volume in the right lower corner. If you look at the red line there, it actually doesn't start at the beginning. It starts somewhere at a follow up time point. The volume of that lesion, if you add it to the green line which is the volume of the lesions from baseline, leads to still a shrinkage of the total
[01:34:30] tumor volume or a reduction of the total tumor volume over time, meeting the criteria of a response, but now factoring in a new lesion. The new lesion itself, it may just be the immune system recognizing micrometatstatic tumor cells, an actually desired drug effect, or an initial growth of a tumor before it shrinks. In either case, if it's put together with the rest of the lesions, it can still define a response.

Here is then the question that everybody wants to answer: Are these new
[01:35:00] observations in any form or shape related to better outcome of the patient, meaning the gold standard endpoint of survival? This is retrospective data, so it was not prospectively defined because at the time of the ipilimumab program, when it started, we didn't know what we would expect or what we should expect. In other words, we had to do this retrospectively. Many new analyses have actually been done prospectively. It shows you here that the conventional responses, which
[01:35:30] are the top curve, have the best survival outcome. The progressive disease patients that are progressive using new criteria and old WHO criteria, they have the worst outcome. The line in the middle is new. That line represents the three new patterns that I've shown you. There's clearly a difference in survival outcome, and if that carries forward, it actually would make a case for why measuring these new patterns can be meaningful.

[01:36:00] I'm adding now another graph just to show that there has actually been progress from the ipilimumab early days to today. This is from the pembrolizumab program that was shown at ASCO in 2014. It was one of the first graphs that came out after ipilimumab. It shows the same phenomenon. There is a variety of new patterns, and those new patterns do better on survival than that is the case with conventional progressive disease.

[01:36:30] Then a few other observations that are worthwhile, and we somewhat heard that already from the FDA and actually from Larry Schwartz as well, is the kinetics that you see on survival and how do the non-survival endpoints response or progression-free survival ultimately relate to the survival outcome. We know historically from cytotoxic agents or from targeted therapies that there is often a dichotomy between the PFS and the OS endpoint. Often the hazard ratios for PFS
[01:37:00] or the benefit on PFS has been greater than that what you see for survival. With immunotherapy, we're making the opposite observation. Let's have a look at that.

This is ipilimumab. Here you see the late separation of curves. You see the plateau of the curve which shows long-term survival for a sub-group of patients. When you look at the response rate, it's 11%. Look at disease control, it's almost 30%. The long-term survival at two years as it's shown here that was one of the secondary
[01:37:30] endpoints measured, is actually 24%. There is a gap here between response and survival. If you add the stable disease patients into the mix, which disease control rate does, it helps to bridge the gap. The first indicator, you need to look at this slightly differently.

Here let's look at another example where it's not exactly the same. For the PD-1 blocking agents, we have seen that response can be high, not in every disease but
[01:38:00] pretty well across the board, and that PFS can be quite favorable. The PFS and the OS outcomes can be very similar. That uses standard criteria. The PF OS relationship is almost 1:1. We see that in a variety of settings. It's very different from ipilimumab.

Then of course, Renzo made this case already for Provenge or sipuleucel-T, that is a
[01:38:30] cell-based cancer vaccine, that has shown no benefit on PFS at all. Basically no responses, but a reasonable-sized benefit on survival. That falls more in the category of cancer vaccines where you historically have not seen responses. We're still grappling with how we make cancer vaccines work, but there is still a lot of interest in the field. If we now think at the spectrum of immunotherapy agents that are being investigated, they're not all checkpoint modulators. They're not all PD-1s.
[01:39:00] You might find within the checkpoint category, we're going from blocking agents like PD-1, PD-L1, to antagonizing agents like OX40, ICOS and other ones, and you see new dynamics. For us to be able to measure what these agents do is important. I believe for that you will need a more flexible tool than conventional response. For PD-1, you may not, but for the spectrum of immunotherapies we're dealing with,
[01:39:30] that flexibility might have to increase.

I'm drawing now the conclusions from the data that I've shown you. The unusual response patterns and delayed effects reflect the underlying biology of the immune system. This is something we thought at the time of ipilimumab, so six, seven years ago, novel patterns of response or detectable disease represent clinical benefit because they do actually correlate with better survival. Conventional
[01:40:00] chemotherapy-driven response or PFS does not account for all benefit patterns and therefore und-

Section 10 of 45 [01:30:00 - 01:40:04]

Section 11 of 45 [01:40:00 - 01:50:04] *(NOTE: speaker names may be different in each section)*

Axel Hoos: ... Forms of PFS does not account for all benefit patterns and therefore underestimates benefit. That might vary, based on what I just said, between different types of immunotherapy. Many evaluation of [inaudible 01:40:14] can be impacted by the late clinical effects due to non-proportion hazards, it may require modified statistics.

[01:40:30] Think about this as something that's 6 years old. If you now take this forward, what are the clinical trial challenges that result from that, and only looking at old endpoints, not the new ones yet. Conventional overall response rate of PSF underestimates the effect and what could that lead to? It could prematurely discontinue a new therapy, especially if it's a therapy that doesn't deliver a higher response rate, but could otherwise potentially improve survival.

[01:41:00] The potential clinical trial failure might occur if a conventional effect is insufficient to meet the study objectives. Let's say you want a 20 percent response rate, you get 10 percent conventional response and 20 percent unconventional response. Then, if you measure the unconventional response you might meet the objective. If you don't, you would not meet the objective. Of course this is widely variable, there's a variety of things we need to consider, but the concept makes sense. Having the tool that enables you to measure that should actually enable us to be better in immunotherapy investigation.

[01:41:30] Premature treatment discontinuation due to conventional progression is a problem we've all faced with many patients, where you actually think, "This patient is benefiting from the therapy," but the measurement on the CAT SCAN tells me I actually have to stop treating now because it is conventional progression. That's something we need to overcome. We have started doing that. I'll speak to that in a [01:42:00] second. That's from a clinical trial perspective and from a treater's perspective something very relevant for clinical practice.

Then there's another aspect here, when you're actually sequencing therapies. We're now in the era of combination therapy where this will become a much larger question. If you're sequencing therapies it could be difficult to actually discern the effect of a previous therapy, from that of a following therapy, depending upon how you determine progression in between.

[01:42:30] Let's look at how can we improve these endpoints. There's a bit of history here, you've spent more than a decade to start this process so I started a timeline. Beginning in past century as [inaudible 01:42:43] clinical observations have been made, a lot of them, and published around unconventional and mixed-response observations. Some lesions get big, some lesions shrink. Very heterogeneous [01:43:00] pictures. Nobody knew how to give credit to that, or what that means, ultimately it got dismissed. That was more than a decade of publications, mostly with cancer vaccines that hasn't really gotten much recognition. Now, what did we do? We started doing some workshops with the Cancer Immunotherapy Consortium and started describing the phenomena and get community buy-in to managing these new phenomena. One observation was the delayed separation of capillary markers. [01:43:30] This was an analysis of all Phase 3 trials that have even been done with immunotherapy and were in a public domain. We did that in 2006. We came up with the conclusion, those that has a separation of curves, they have a delay. What does this mean? Do we actually need to do new trials differently. That was very important.

Then we came with the immuno[inaudible 01:43:47] response criteria that was actually done jointly between CIC and BMS using a [inaudible 01:43:53] that was
[01:44:00] published in 2009. From there we now actually have seen nice expansion on that activity going into new versions of the original IRRC. Going getting more nuanced, considering more data. They are labeled as IR-resist. The new version of I-resist, the IM-resist, and the IR-lugano. I just had to put that in here, which is for lymphoma. It was just published in blood and it uses the same basic principles, but it moves on from solid tumors to hematologic malignancies.

[01:44:30] Then of course, besides advancing criteria, we're also getting more data. I hope the next session will actually show us a lot of that is a lot of new data that enables us to say there's some solidity to these observations.

Here we are today, hopefully this will propel us to the next level after having spent the better part of the decade trying to get to this point.

[01:45:00] I say a few more things about the IR response criteria. As we heard from Larry Schwartz, I think he said it very well, IR response criteria were not designed to give a very, very specific way of managing this problem. They were concepts that you can apply in a variety of ways. We of course, in order to use them in a clinical trial, had to apply them in a very specific way, so we used WHO as the under-pinning system. The reason was the [inaudible 01:45:24] program had been using WHO criteria for all their trials. It made no sense to try and deviate and move to resist.
[01:45:30] We used WHO but we defined principles that could easily be translated to resist, and that has subsequently happened. Confirmation of progression was a component. Measuring new lesions was a component. Total tumor burden, including new lesions, that means measure the lesions and add them to the tumor volume. Treat post initial progression, not indefinitely, but at least that first time when progression occurs, allowed the patient to still benefit from therapy until you
[01:46:00] have really confirmed progression.

Then of course, if you take those components, you define response criteria, and apply them to the endpoints that we're interested in you would then create things like IR response, or IROR, [IIDSCA 01:46:14] including stable disease and IIPFS, which might actually become the endpoint theta we are the most interested in as we progress this story.

We were then asked by clinical cancer research when we published this the first
[01:46:30] time around, about 6 years later to give a progress update. That was published last year. It makes a few nice points. The first one is, we actually have seen transferability of IIRC concepts from WHO to resist. Almost everything we are talking about now is resist. The application f the IIRC has been done beyond melanoma. It's true most of the data was melanoma initially, but we have now seen it in lung cancer, we have now seen it in other diseases. The inclusion of the
[01:47:00] concept has been put into regulatory guidance documents for exploratory purposes so far, but that is useful. The expansion of IIRC into more concrete criteria has happened. That's what I've just listed for you. Then of course, we have seen a class

effect. Once you apply this broadly, different immunotherapies can show these patterns. It's not just a [inaudible 01:47:19] it's not just a one-drug phenomena.

[01:47:30] This is what I view as worth defining progress in the last 5-6 years. Here, if you look at the different adjustments of IIRC we have seen, IR-resist, I-resist, IM-resist, and IR-lugano, they have a lot of overlap. The original principles that I just named, they're all reflected in almost all of the criteria. There are a few nuances difference. The one that I would point out to is the IM-resist, is actually aiming to look at progression post the confirmed progression, or response post the confirmed
[01:48:00] progression, which is expanding our concept a little bit further because there are very late effects that you might sometimes see that would certainly not be captured by most of the variations that we see here.

[01:48:30] I mentioned already in some regulatory guidances we have seen inclusion of the basic principles of IIRC. This is a cancer vaccine that the FDA had issued. This is also by now 5 years old. There is an EMA guidance which is in the process of being updated right now where the same principles have been included. We actually took a workshop of the CIC in 2014 that was 5 years after publishing the criteria to see how much further have we come. That was a bit premature compared to where we are today, but it already points us in the right direction. There is a class effect here across a variety of different immunotherapies and a class effect legitimizes the use
[01:49:00] of criteria much more so than if you look at a single agent.

I come to the survivor question. The point has been made already; if survivor gets better from any new therapy that is included in our [mememtario, 01:49:15] then the tail of the curve will go up, more patients will live longer, and any new therapy that you will introduce will have a much harder time to do a meaningful trial with a reasonable duration on a conventional survival endpoint.

[01:49:30] We got to find a way to make this more palatable. Here is a slide that I just love to show. It's an [Epineveale, 01:49:35] Phase 1 follow-up showing a very high plateau of the survival curve in metastatic melanoma. This data still needs more follow-up and more Phase 3 extension, but it is exciting to see combinations can produce benefit and we are achieving this increase in survival for patients.

[01:50:00] Now let's have a quick look at what are the challenges and solutions for survival, and you will hear a lot more about this when the statisticians speak later ...

Section 11 of 45 [01:40:00 - 01:50:04]

Section 12 of 45 [01:50:00 - 02:00:04] *(NOTE: speaker names may be different in each section)*

Axel Hoos: And you will hear a lot more about this when the statisticians speak later. The challenge is obviously the non-proportion hazard. In order to solve that we need new statistical models, potentially in some instances it had been proposed to over-power studies. Because you can draw then from that extra power when you encounter the late effect. And that seems to dilute your original power. And then
[01:50:30] you can deal with sensitivity analysis. Which help understand things but not necessarily saving a primary endpoint in a trial.

To increase duration of trials with continuously improving survival, I made that point, the best way to deal with this is to use an earlier survival read-out. Milestone survival has been proposed. I think Tai-Tsang Chen will speak about this later in much more depth. And of course alternative PFS endpoints. Here, IPFS, IMPFS, or IRPFS, whatever it might end up being, could make a big difference. If it's a credible
[01:51:00] endpoint that reflects the biology it could actually really make a difference.

And then the same applies actually for the other two points I'm making here. The dilution of an effect of other survival influencing therapies, post-progression, can also be somewhat circumvented if you use new endpoints. The same is true for the crossover problem that might occur if you have trials where survival impact of
[01:51:30] different agents is great, and you're comparing them against each other, either alone or in different regiments, it could cause a problem for crossover. Patients might when they progress they might want the other therapy. And therefore, having earlier readouts either on survival or on PFS preferably, will help this.

So with that said, the ultimate goal is that we should follow, I think we're all aligned on this in principal, is that we need better tools, which would be immuno therapy adapted endpoints. And they're not excluding the use of the existing tools, they're
[01:52:00] just additional. Then we harmonize the criteria so we pick one. After we have defined which one is actually the most favorable which response criteria should be used, we pick one and make that the community preferred approach. We need to reflect the biology in this, obviously. We want to accurately capture medical benefit, that's where it all began, because we think we're missing something. And
[01:52:30] then ultimately, we want to enable future trials where the survival is continuously increasing, and we want to do that with regular to incredibity. That means the endpoints we choose should be endpoints that we all can agree to, as meaningful and therefore potentially primary endpoints for pivotal trials. We're not there yet obviously, but I'm hoping that today will get us one step closer.

So I'm closing with this, and I think Lorenzo will lead us towards a panel discussion. Thank you.

[01:53:00]

Renzo Canetta: [01:54:30] We have postponed the panel discussion until after the presentation of Dr. Latimer so we're going to take the break now, and resume at 10:15 sharp. Thank you.

In order to keep with the time I would ask you to come to your seats. We need to resume. The last of the talks were dedicated to the traditional endpoints. And before the panel discussion that we will have at the end of the talk, is by Dr.
[01:55:00] Nicholas Latimer from the University of Sheffield in England. Nick will talk about the impact of non-proportional hazards on long-term follow up curves. Nick.

[01:55:30]

Nicholas Latimer: Okay, thanks very much. I wanted to thank the organizers very much for inviting me to come along to this. It's been a really interesting couple of days for me so far. But

I think my talk might be a little bit different to what we've had so far. Hopefully it will still be interesting. But the reason it might be different is because I'm a health economist. I'm not a clinician. I'm not a statistician. And although most of my
[01:56:00] research is on survival analysis, it's in the context of what we need to conduct economic evaluation. So that influences the types of outcomes that I'm interested in. And that will probably become clear as I go through my slides.

I'm going to first give a little bit of background on the standard survival analysis
[01:56:30] techniques. And I won't spend long on this because I think most people here will be well aware of them. I'm going to spend a little bit of time talking about parametric survival analysis, which is what we absolutely use a lot of in the economic evaluation. Before moving on to what the particular challenges, presented by immuno-oncology are, and how we might go about addressing them, I think it's worth noting from the outset that today I'm just talking about modeling survival
[01:57:00] data, the data that we absolutely observe in a trial and beyond it. But I'm ignoring problems associated with things like treatment switching, which were alluded to earlier, where patients in the control group maybe switch onto the experimental treatment. That's actually what most of my research is on, methods to adjust for that kind of switching. So if you are interested in that then feel free to ask questions or come and find me. But I'm not going to be talking about that in this talk.

[01:57:30] So first of all standard survival analysis. So I don't need to spend long on this because as I said I think everybody's well aware of these things. But just it's useful to have a little bit of context at the start. So survival analysis is simply the analysis of time to event data from a specified time origin, which in a trial is usually the point of randomization, until the occurrence of a particular event or endpoint which would be anything. Usually we're looking at disease progression or death.
[01:58:00] And it is very commonly used in evaluating the effectiveness of health care interventions. And we usually see things like Kaplan-Meier curves, we're presented with statistics like median survival, and hazard ratios. So this is an example of a Kaplan-Meier curve which tells us the probability of surviving until time t. We can use Kaplan-Meier curves to estimate medians, and means restricted to the time period that we've got data for. And we can use log rank tests to compare the
[01:58:30] curves statistically to see if there's a statistically significant difference between those survival curves.

If we want to actually quantify the treatment effect, then we can use a cross-proportional hazard progression model which is probably the most commonly used model for this purpose, for producing a hazard ratio. And the hazard ratio simply compares hazard rate between treatment groups and make the pivotal assumption
[01:59:00] that there are proportional hazards. So the baseline hazard rate can change over time but the hazard ratio cannot.

So I wanted to highlight a couple of issues with these standard, non-parametrical, semi-parametric, survival analysis techniques. The first is if the pivotal assumptions associated with these techniques don't hold, then possibly these analyses are not

[01:59:30] very useful for us. So the log rank test and hazard ratios from Cox models are reliant on the proportional hazards assumption. If that assumption doesn't hold then those statistics possibly aren't valid for us.

The second issue that I wanted to highlight is even if those assumptions did hold, are they really telling us what we want to know about the effectiveness of the new
[02:00:00] treatment? So from a health economist's perspective, we conduct economic evaluation with the primary goal t-

Section 12 of 45 [01:50:00 - 02:00:04]

Section 13 of 45 [02:00:00 - 02:10:04] *(NOTE: speaker names may be different in each section)*

Nicholas Latimer: We conduct economic evaluation with the primary goal to inform decisions made about allocating health care resources and what we want to do is maximize the health that we're getting from a scarce healthcare budget. What we really want to do is estimate the effectiveness of providing a new treatment for an entire disease population.

[02:00:30] The survival input that goes into our economic model usually isn't a hazard ratio and it's not a median survival time, it's mean survival. We want to know what's the mean survival time and what's the mean survival benefit associated with the new treatment?

If we've got a clinical trial with censoring, which we usually do. We never have complete follow-up where everybody has experienced the event we're interested
[02:01:00] in. If we estimate mean survival based purely on the trial period, then it's going to be a restricted mean. It's going to be a downward biased estimate of the total mean overall survival.

What we really need to do and what we do for economic evaluation commonly is fit a model to this data and extrapolate beyond the trial period so that we can estimate total mean survival times. If we want to do that and there's a role for
[02:01:30] parametric survival analysis instead of the more standard non parametric survival analysis. This is as I said, what we commonly do for health technology assessment in the UK and in other areas where we're trying to estimate the cost effectiveness of new drugs. It's not only mean survival that's important. We also take into account quality of life so we actually look at the cost of quality adjusted life gained
[02:02:00] associated with a new treatment. A fundamental part of that analysis is what is the mean survival benefit associated with the new treatment?

Parametric survival analysis and I'm just going to give a really brief overview of this. The reason that parametric survival models are useful for us is because they assume that the survival data follows a specific underlying probability distribution and that allows us to project survival beyond the end of the trial. There are several
[02:02:30] different models that we can use for this which causes some of the problems associated with undertaking survival analysis. Really, the key is we need to try to identify the most appropriate model on a case by case basis so that we get a credible projection of survival.

Just a quick overview of some of the standard models that are available to us, which have commonly been used in health technology assessment in the past. The
[02:03:00] simplest parametric model is the exponential model, which assumes that hazard rates stay constant over time. That's obviously seen as often not a valid assumption to make because it's quite rare that a hazard of an event that we're interested in, something like progression or death will stay constant over time.

Although exponential models have been used in the past in economic evaluations,
[02:03:30] generally I think there's a move away from those and they're not used very often anymore. [Wible 02:03:31] models and Gompertz models are a little bit more flexible. They can represent hazard functions so hazard rates that can increase or decrease monotonically over time. You know that long after you can assume that the hazard rates stay constant. You have to assume that the hazard rate will either increase or decrease monotonically. If you're going to use one of those models,
[02:04:00] that's the key assumption that you have to consider whether it's valid or not and often it may not be.

Logistic and lognormal models are again a little bit different. They can represent hazard functions which allow an initially increasing hazard followed by a decreasing hazard in the longer term. Often these types of models are characterized by long tails because you have this decreasing hazard in the long term. The problem that you sometimes get with those is that in the very long term, the hazards that they
[02:04:30] are projecting fall lower than the hazards observed in the general population, which is seen as something that's not credible and so the model needs to be adapted to address that.

Probably the final one of the standard parametric models that are used fairly commonly, although this particular one isn't used very often is a generalized gamma model, which has 3 parameters within it which allows it to be a little bit more flexible than those other models I've mentioned. One of the useful things
[02:05:00] about the generalized gamma model is that it includes the exponential, the Wible and lognormal models is special cases within it. If you fit the generalized gamma model, if in fact, one of those simpler models does represent the data well. It will effectively reduce to one of those models but it gives you the added flexibility so that if one of those models aren't appropriate, it will provide you with a better fit. Still, it may not be flexible enough to represent some very complex hazard
[02:05:30] functions that we might observe.

A couple of issues that I wanted to highlight with parametric survival models, they can be fitted assuming proportional hazards and proportional treatment effect. You get 2 different types of parametric model, either proportional hazards model or accelerated failure time models. With either one of them, you can fit them assuming a proportional treatment effect.

[02:06:00] As for the Cox model, if in fact we don't have a proportional treatment effect, then doing that will not be appropriate and won't give us a good representation of the

data that we've got. We can get around that with parametric model fittings so we could, for example, fit a set of parametric models to each treatment arm. For example, we could fit one Wible model to the control group, survival data and a different Wible model or even a different distribution entirely if we think that the
[02:06:30] treatment totally changes the survival distribution to the experimental group and then we can fit those models, project survival and not assume proportional hazards and still come up with our estimate of the mean survival benefit.

Really our kind of summary statistic of the treatment affect is mean survival benefit rather than something like a hazard ratio of median survival.

This graph just shows lots of different parametric models fitted to some data. It's
[02:07:00] just to show that the different models give significantly different curves. That's because they all assume the data follows slightly different distributions. This means that the model choice is absolutely critical. Often we see in the technology appraisals in the UK, that the choice of parametric model can completely change your decision as to whether a drug is cost effective or not. It's often the most influential parameter within the economic model.

[02:07:30] It's really critical that a systematic approach is taken to identify the model that is most appropriate for your particular case. The way that that's generally recommended to be done is that you need to identify a model that fits your observed data well and something which we call internal validity, but also represents a good projection of what you expect the hazards to be in the longer term beyond the trial period. That's called external validity and you need to look at
[02:08:00] data other than just the trial data to inform that. It may be clinical expert opinion and what you expect to happen. It could be longer term data from earlier phase trials which gives you more information about the survival in responders for example. It could be data from registries. You need to come up with some justification as to why the projection that you've come up with is credible.

[02:08:30] So far I've kind of talked about standard non parametric survival analysis and the parametric survival analysis that we use often in economic evaluation. What about immuno-oncology? What's different about immuno-oncology that changes or may change the way that we need to do this survival analysis? I think it's really important to point out that survival modeling in any disease area for any drug is
[02:09:00] very rarely straightforward. It's not uncommon at all for us to analyze trial data and decide that the proportional hazards assumption is not valid. Often we have very limited long term survival data so the projected part of the survival curves is very important. Sometimes we get hazard functions that do appear to not follow one of the standard parametric survival distributions. These things happen. They have
[02:09:30] happened many times over the years and we need to think about ways to deal with them on a case by case basis.

This really isn't specific to immuno-oncology. It's a common issue. Nevertheless, immuno-oncology at the moment does appear to be characterized by certain features which do have implications for how we would do the survival modeling.

[02:10:00] We can certainly talk more about those. There are very likely to be many other features, but the 2 that I've kind of highlighted here are firstly, immuno drugs may ...

Section 13 of 45 [02:00:00 - 02:10:04]

Section 14 of 45 [02:10:00 - 02:20:04] *(NOTE: speaker names may be different in each section)*

Nicholas Latimer: ...that I've highlighted here, rather. Firstly, immunodrugs may not exhibit a proportional treatment effect. Secondly, immunodrugs may be associated with a delayed effect, long-term survivors, and therefore complex hazard functions. Those are two challenges. How might we deal with those?

[02:10:30] The first one: immunodrugs may not exhibit a proportional treatment effect. These figures here are from a paper by a speaker in one of the later sessions today, [sai-sun 02:10:29] Chen, and the full reference for the paper is given in my references slide at the end. What we can see here, the top figure shows the more classic case that represents proportional hazards, where you've got two survival curves: one for the control group, one for the experimental group, which diverge early on in the trial and stay apart with proportional hazards.

[02:11:00] The graph at the bottom is what we may expect to see, or what we have seen in some cases, for immuno-oncology drugs, where we've got this delayed clinical effect. Early on there is no difference in survival. In the longer term, the curves begin to diverge. That's classically against the proportional hazards assumption. In these situations, Cox models, hazard ratios from Cox models, are probably not appropriate. Log rank test results are probably not appropriate, and any parametric model that assumes proportional hazards or a proportional treatment effect may [02:11:30] not be appropriate either. So what can we do?

[02:12:00] We can fit Cox models with time-dependent treatment effects, but one of the key advantages of a Cox model is that it gives us this simple, single hazard ratio, which is easily interpretable as a summary of the survival data. If we incorporate time-dependent treatment effects, you no longer get something simple like that. As I've pointed out, from my perspective, from the perspective of somebody doing an economic evaluation and deciding whether we should reimburse a drug, then shouldn't we be more interested in mean survival anyway, instead of a hazard ratio? Hence the non-parametric analysis might not be that useful.

[02:12:30] As I said, parametric models could be fitted independently to each treatment arm in this situation, which means that we don't need to assume that there are proportional hazards, and it may allow us to get a better estimate of the survival [gain 02:12:29]. This is something that has been done regularly in technology appraisals in the UK, where models have been fitted without assuming proportional hazards and incorporated within the economic evaluation.

[02:13:00] The second issue: complex hazard functions. As we've seen with lots of the graphs that have been shown in earlier talks, immunodrugs may be associated with a delayed effect, long term survivors, and therefore a complex hazard function. This

diagram here is some made up simulated data that I came up with, which shows a hazard function over time, a smoothed hazard function over time. Early on, the hazard is quite low. You may see that in some trials, because eligibility criteria mean that patients are actually relatively fit in some circumstances when they enter a trial, so the hazard of them dying very quickly might be quite low. But often
[02:13:30] they have serious diseases, and so the hazard increases in the shorter term of dying. Then in the longer term, you're left with people who have done well. They've got longer term survival, they may have responded to treatment, they've got better prognoses. So you may see a hazard that begins to fall in the longer term. In fact, you may see that even in the longer term than that, as age-related mortality becomes more important, that hazard might begin to increase again, in the very
[02:14:00] long term.

This hazard function has a turning point in it. We're seeing that the standard parametric models will struggle to deal with this. What can we do about that? We could use flexible parametric models. Flexible parametric models were first introduced by Patrick Roysten and Matt Palmer in 2002, and there's been lots of
[02:14:30] papers about them since then. Particularly in recent times, researchers at the University of Leicester have written a lot about these and developed very useful software packages to allow them to be fitted. What they do is use restrictive [cubics blinds 02:14:41] to estimate the shape of the [log umes 02:14:43] of hazard function, and allows them to accurately reflect complex hazard functions with turning points in them.

[02:15:00] This is an example. On the left hand side of this figure, we've got a simple [weibel 02:15:01] model fitted to this complex hazard function that I showed you before. As you can see, and as we talked about, the [weibel 02:15:08] model can only deal with monotonic hazards. In this case it's estimated a monotonically increasing hazard over time, which doesn't allow you to reflect that turning point in the data, and means that you end up getting a model that doesn't fit the data very well. It doesn't have that plateau in the survival curve in the longer term.

[02:15:30] On the right, we've got a flexible parametric model. I think I incorporated [false blinds 02:15:33] in that model. You can see that that does allow you to reflect that turning point in the hazard function. If I'd incorporated [morse blinds 02:15:41] in the model, it would have been able to follow that even more closely. You can see that does allow you to predict the survival function much more closely, and it does have that plateau in the longer term. This is something that we could use, if we've got these complicated hazard functions.

[02:16:00] There's other alternatives, as well. It was mentioned earlier that there are cure models, parametric cure models. Sometimes it might appear that a proportion of patients have actually been cured. We can use a parametric cure model to estimate the cure fraction, and then you essentially model the survival experience of cured patients and uncured patients with different distributions. A slightly weaker
[02:16:30] assumption would be that we don't have to assume that there's a cure. We could just assume that there is a distinctly different survival distributions amongst

different groups within the data. Then we could just use a parametric mixture model rather a cure model, which models the data as though there are distinct distributions within the data. Possibly different groups of patients within the data with different survival experiences.

[02:17:00] This is an example of a mixture model, fitted to that complicated data that I showed. Again, we've got this [impo-aigo 02:17:06] model on the left that doesn't fit the data very well, and on the right we've got a mixture [weibel 02:17:11] model fitted to the data. Again you can see that because we've now got a combination of two [weibel 02:17:18] models being used to represent the data, we can reflect this turning point in the hazard function. We can get this plateau in the survival curve in
[02:17:30] the longer term. This could be something that could be useful for us if we think we've got these kind of turning points in hazard functions.

It's important to point out that these techniques, as with most techniques, most modeling techniques, most statistical methods, have important assumptions and limitations associated with them. For flexible parametric models, it's important to note that they extrapolate beyond the data, basically using the final segment of the
[02:18:00] data. That may or may not provide us with an accurate projection of survival over time. We really need to be careful to see whether we think the projected part of the curve is valid and credible. The cure or mixture model is maybe, some people argue, have got biologically a preferable basis, as they don't involve segmenting the survival curve, but they have an underlying belief that there are different
[02:18:30] distributions of patients within the data, which maybe fits the kind of ideas that I think we might have about the survival experience on immuno-oncology drugs.

But, again, there are issues with this. Can we actually prove that the assumption of a cure is reasonable, for example. Sometimes if we don't have much survival data, the model will struggle to accurately estimate what the cure fraction is. If we use a
[02:19:00] mixture model, can we justify the assumption that there are distinct distributions of survival within patients in the trial? We still have the issue that we need to select appropriate parametric models. Within the mixture model, do we use a [weibel 02:19:12] model? Do we use Gompertz, do we use Log-normal, et cetera. We can't just plug these methods in and expect them to immediately give us sensible results. We need to think carefully about how we do it, and justify the choices that we
[02:19:30] make, because there are many choices to make.

A few conclusions. Immuno-oncology may present non-standard issues for survival analysis. I say "may" there, because we do have long-term data for some drugs, but not that many. So, we might not necessarily know what is going to happen to overall survival in longer term with some of these treatments. If they do have these non-standard issues, then it has important implications for the standard methods
[02:20:00] of survival analysis that have been used, particularly in regulatory-

Section 14 of 45 [02:10:00 - 02:20:04]

Section 15 of 45 [02:20:00 - 02:30:04] *(NOTE: speaker names may be different in each section)*

Nicholas Latimer: Analysis that have been used, particularly in regulatory submissions in the past. We

may need models that allow time to [inaudible 02:20:10] an effect, so we may need to model survival curves independently, and flexible parametric models and cure models and mixture models all offer potential solutions for doing this. Each of them requires careful justification, and as with any survival analysis exercise, we need to
[02:20:30] be very careful about demonstrating that the model that we've chosen is an appropriate one and comes up with a credible projection of survival over time if we're trying to estimate mean survival.

One further conclusion that I haven't written down there that I thought is relevant to point out is that because of the outcomes that we use in economic evaluation, the amino oncology issues of non-proportional hazards doesn't necessarily cause us
[02:21:00] as much of a problem, or cause us to need to change the techniques that we use as it may do for maybe regulators who have previously relied on hazard ratios. We're already using parametric modeling commonly, and commonly seeing or deciding that it's not necessary or appropriate to assume proportional hazards. We're already using some methods to help us deal with that.

These are just some references to the papers that I referred to as I went through the presentation and that's all. Thanks.

[02:21:30]

Renzo Canetta: Thank you [inaudible 02:21:30]. I'm inviting the speaker to sit at the podium for the panel and I'm announcing that Dr. Shenghui Tang from the FDA is also joining the panel. As the speaker are taking their seat, let me start with an anecdote. Back
[02:22:00] when resist version 1.1 was created my former company and many other companies donated their databases for the work that the independent resist committee did. Because of that we were given the courtesy to review the paper before the publication, and I still remember that our first comment was, we live in the era where TKI are coming to the scene, novel therapies are coming to the
[02:22:30] scene, and we may actually be more focused on time to event outcomes rather than just tumor shrinkage.

I'm saying this not to say that we are good, because that's not important. I wanted to make two points. First, that it's possible to share data, so this is an example where we have done it with several other companies. Second that we always
[02:23:00] should think about the future, what the future brings. With this introduction we start the discussion.

I'll start with a question concerning objective responses. Our role as investigators is to start from the existing knowledge and challenge it, verify it, or advance it. Yesterday Suzanne Topalian spoke about neo adjuvant trial design. Obviously
[02:23:30] responses are going to be extremely important for neo adjuvant trials. One of the old tenets is that in immunology takes a long time to mount a response. That will obviously have some impact if true, in the neo-adjuvant design. I would like to hear the opinion of the panelist about this aspect. Time to response as this pertains to
[02:24:00] study design. Whoever wants to take it.

Of course it takes a moment to process intelligent questions from you. Time to response is a key factor, we've learned a lot about this in the recent years. For neo-adjuvant trials, I don't think we have actually spent, we have generated any data on
[02:24:30] this yet, to my knowledge. I think of course time is a factor, so the longer this will take the harder it will be to implement it in an intelligent trial design, because the time to surgery is a key factor.

It depends on the setting. If you can induce a immune response and induce a pathological response if you want, in the tumor before you resect it, there's enough time for that to happen, then this is a very important thing to do. I'm not sure yet
[02:25:00] how that will play. Depends on the individual [selling 02:25:02].

I would only add that this is perhaps the value of the introduction of the spider grams. When you look at spider grams that have been generated for several checkpoint inhibitors that actually indicate, and this is particularly true for combination therapy, that the time of response can be actually be dramatic and very short, so it would dispel. In addition, you would have the opportunity to
[02:25:30] expose tumors in the adjuvant setting to immunological priming so to speak, and that will have an effect even after post-surgery.

Dr. Tang?

Shenghui Tang: I want to have a couple comments on in general response rates you said in the trial. As we already know there is a DNA factor and particularly compared with
[02:26:00] chemotherapy and we can get the response very quickly but when we look at the immunotherapy trial. Of course most of the response were occurred about 6, 8 weeks but you also see much later responders so if, so when we look at the response we need to look, think about how long the follow up will be and to not
[02:26:30] underestimate the response rate.

Renzo Canetta: Thank you. We open it to the audience. Oh to Dr. [Ribas 02:26:37] first because he rose before.

Speaker 3: Quick comment on that. For the economic valuation perspective. The time to response could be very important if it's related to quality of life as well. If once you've responded you have a higher quality of life, then we would really want to
[02:27:00] take that into account in evaluating the benefit of the new drug.

Speaker 2: Okay, as a reminder for the web audience, please introduce yourself even if we all know you.

Speaker 4: My name is [inaudible 02:27:09] Aribas, I'm a medical oncologist from UCLA. I want to first thank the panelist for the great presentations, but I want to challenge the concept of slow progression. We heard it be talked about, slow progression and tumor flare. I think there's very little evidence that it's mechanistically based to
[02:27:30] slow the progression. I cordially disagree with my friend Axel who's on the images that you showed showing that that's an inflamed tumor. When I've seen it, I've

seen true progression followed by a delayed response. To me is a delayed immune response that goes through a period of large tumor. The images I've seen from others seem to suggest to me that that's the case. Whenever we've done a biopsy
[02:28:00] we've seen that it's mostly tumor with very little infiltrate. Eventually it's infiltrated and the tumor responds. The majority of tumors who respond to these immunotherapies have a distal infiltrate with very few inflammatory cells and you don't see inflammation, you see the tumor get smaller over time.

This is a true phenomenon but it's a rare phenomenon. It's around 1 in 20 patients. That's known from the perspective testing of this. We're trying to repeat the
[02:28:30] evaluation of large studies based on 1 in 20. Whenever it's been tested side by side the [Pembro 02:28:39] phase 1 trial had the side by side testing of [reces 02:28:42] and IRRC and its published and the response rates when you evaluate both of the perspective was very similar, telling us that the effect is very small. We have the clear examples, but the overall effect is small. I'm trying to say is that the majority,
[02:29:00] the difference between evaluating immune oncology agents may be less dependent on pseudo-progression and more dependent on other factors. Pseudo-progression, I think it's a misnomer.

Speaker 2: Okay so I answer this of course as I started the whole thing. I agree with you actually in principal that it was not meant to say that there's only lymphocytic
[02:29:30] infiltration in a tumor that can make a tumor look bigger. Of course it can be regular progression that then later is followed by an immune response that can shrink it. I agree that pseudo-progression, call it a misnomer or not, actually entails both phenomena. How they distribute, I don't really know that. We haven't biopsied enough tumors to really know. But clearly both of them exist. As it comes to the frequency now, the [inaudible 02:29:55] data that I have seen and the one
[02:30:00] that I showed actually had about 16% of patients in that about 500 patient data set

Section 15 of 45 [02:20:00 - 02:30:04]

Section 16 of 45 [02:30:00 - 02:40:04] *(NOTE: speaker names may be different in each section)*

Speaker 1: Patients in that about 500-patient dataset that had a phenomenon that was either a delayed response or stable disease following an initial-

Speaker 2: (Russian accent) Stabled disease, yeah ... That's what I'm saying. Progress ... Increasing size plus [inaudible 02:30:17] response was around 5 percent, 4 percent.

Speaker 1: Yeah, the key thing, though, is that benefit, stable disease falls in the same pocket. If you can't call it stable disease and count it as benefit because there was this
[02:30:30] initial blip of tumor growth, it still takes a lot of patients away from the analysis. The way we've done this from the beginning is, we called it "immune-related disease control," which includes the stable disease patients. You're right, it's a small number of patients that actually have a true delayed response.

Speaker 2: Actually, I want to point one thing, that if it wasn't for your and [Jet's 02:30:54] contribution to these phenomenon, we wouldn't have had a confirmation of
[02:31:00] progression which has allowed us to study this phenomenon with RECIST

evaluation. Now we know that it is true. It's reproducible, but it also happens in the control arm.

I think there was a BMS trial on it. Maybe [Rensa 02:31:13] knows it, or you know it, where IRRC was tested in the ipilimumab arm and in the control arm, and it was significant number of IRRC in, I think, it was chemotherapy arm. I forgot what it was.

[02:31:30]

Speaker 1: I say one last thing. I try to make that point during my talk. I-O has had a very fast and furious evolution to the great benefit of many patients, but it's still at the beginning. At least I think we are still fairly early in terms of understanding the immune responses and putting new agents into the clinic that will produce new patterns of clinical activity.

[02:32:00] Having said that, what we're trying to do here is really produce tools that help us evaluate what comes next. For the PD-1 blockers, we needed those tools less so than we needed them for ippy or we need them for cancer vaccines, or we may need them for oncolytic viruses. When you actually look at the approved oncolytic virus, T-VEC, that study was actually a response-rate study.

[02:32:30] The approval was given on the basis of a modified response endpoint that included things like treatment beyond progression and growth before shrinkage. You can do that. It was never called "IR response," but the point is the phenomenon has been recognized. You just need to use these tools to understand what's going on.

Dave R.: Can I just make a pathologists' comment on pseudo-progression? I'm Dave Rimm. I'm a pathologist from Yale, and we've actually done 2 autopsies on patients where they had either pseudo-progression or stable disease, and then they died from some other complication.

[02:33:00] In both cases, most of the lesions had no viable tumor, whether they were enlarged or whether they were stable disease. There were multiple lesions in body in these cases, so I think, and perhaps Larry can comment on other methods of assessing to see whether the tumor's viable or whether there's even tumor there. I wonder, and Axel, you made the point as well. Maybe we need to biopsy more of them to see what's really going on.

Male: [Dr. Schwartz 02:33:28]?

[02:33:30]

Male: There, we hope some of our more advanced imaging will help us.

Male: Thanks. There's no doubt that ippy in melanoma probably produces the highest rate of pseudo whatever it is, yet it still is a phenomenon. Frankly, now, I'm concerned about that group of patients. I think I'm also concerned about the group of patients who are truly progressing, but we think that it may be pseudo-

[02:34:00] progression because we've seen some of it already, and we have to distinguish those to be able to understand that both in terms of our trials as well as in clinical care.

Male: Microphone number 1.

Mark Ratain.: Mark Ratain, University of Chicago. I'm a big fan of parametric models in quantitative pharmacology and drug development. I think I want to also point to
[02:34:30] the importance of modeling tumor size. In fact, the first oncology publication in this regard came out of FDA's Office of Clinical Pharmacology, demonstrating the change in tumor size at 8 weeks was predictive at a study level of survival in non-small cell lung cancer.

There have been subsequent publications largely from René Bruno and his group at Pharsight, suggesting that ... This is not in the IO space. That time to tumor growth,
[02:35:00] as modeled across a population in a study is predictive for survival benefit ... In other words, if you have 2 arms, and you can show a difference in time to tumor growth, that predicts for survival benefits.

I think we can use those same tools in the I-O space, obviously with an adjustment for changes in tumor size that are due to inflammation, pseudo-progression or whatever we're calling it. I would really urge industry, FDA and academics to collaborate on such an approach.

[02:35:30]
Male: Thank you. Number 2.

Male: Thank you. Dan [Chin 02:35:36], Genentech and Roche. Thank you to those speakers for providing an incredibly interesting analysis of where we are today around these types of non-classical responses. As you noted, I think a lot of what has been done to date includes looks at our experience with targeted therapies, CTLA-4 targeted therapies and PD-L1, PD-1 targeted therapies.

[02:36:00] In contrast to my good friend Toni Ribas, I think given the 800 cancer immunotherapy combination studies in the clinic today, that, in the very near future, not only are we going to see the continuation of the phenomena that are being described here potentially referred to as pseudo-progression but also non-classical responses. Specifically for pseudo-progression, I think we will be at a point where, in the very near future, not only do we see pseudo-progression, but
[02:36:30] because of our approaches to try to further inflame tumors, we'll actually see nearly immediate and perhaps 100 percent pseudo-progression.

My question to the panel is, "As we build out methods of evaluating non-classical activity for cancer immunotherapy agents, are we trying to develop systems so that
[02:37:00] we can better evaluate CTLA-4 or PD-L1-, PD-1-based therapies, or are we trying to build a system for the future to try to use what we have today to encompass all the possibilities for capturing activity in the rapidly approaching future of cancer

immunotherapy?

Speaker 1: [02:37:30] Well, I would just very clearly state this has to be something that enables us to go into the future. The checkpoints that we're investigating today are the tip of the iceberg. They're exciting. They have made great progress, but we're always looking to the future. We need to be prepared for that, so I would go with your analysis, Dan, and say we will see a lot more of pseudo-progression or delayed response, whatever we call it.

[02:38:00] We need to be able to just investigate this properly. Capture it, capture it in agreement across all the stakeholders in the space, so we can all use the same language, use the same tools. There may be cases where we don't see much of it, and I will acknowledge that, and we can use all the tools if we want. There will be places where we will need those tools to progress new therapies. At the end of the day, that's the ultimate goal, so we have to look to the future.

Female: Can I-

Renzo Canetta: Number 3, please.

Female: Sorry, can I make a comment?

Renzo Canetta: Oh, I'm sorry, go ahead.

Female: [02:38:30] Not just on this but also on the time to response, a comment that was made earlier ... You're having all these new agents come up, and yet you're not ready to figure out what is the right endpoint, so I think it is really leveraging. Like I said, "What data do I have that can better help design or identify endpoints for the future. Then coming up with new endpoints as you are doing the trial is also problematic.

[02:39:00] For example, time to response ... You first off don't know what is that time to response that you are actually seeing in agents, similar population, similar agents. Then what is the threshold that's going to tell you that this is exciting and this is not exciting? There's not a lot historical data, and so that makes it very difficult to intelligently design a trial for the current study but also, trying to think of the future.

[02:39:30] I agree with what was said. I think we have to think about how we are collecting data in a good way now that we can mine and try to maybe ... Maybe we can't take 10 years, but maybe in the next 2 years, based on the data that we have, we are able to say, "These are what we can learn." These are what we cannot learn. These are the type of endpoints. These are the type of analyses." Then maybe, hopefully, for the next agent or the next few agents, we'd have a better answer.

It's very hard to design when you're already bringing in all these agents and then saying, "We have to move together." Almost it's being done in parallel, so you're going to miss the mark in some trials that are ongoing. Hopefully, you'll learn from

[02:40:00] today to better design trials in the future.

Renzo Canetta: Thanks. Number 3 please.

Male: [Gram 02:40:03]-

Section 16 of 45 [02:30:00 - 02:40:04]
Section 17 of 45 [02:40:00 - 02:50:04] *(NOTE: speaker names may be different in each section)*

Female: Trials in the future ...

Renzo Canetta: Thanks. Number 3, please.

Grant Williams: Grant Williams, Philadelphia. As we try to apply these models that Dr. Latimer showed us to the statistical and regulatory design of clinical studies, what are we heading toward? Are we heading toward pre-specifying a certain model or an algorithm of some sort to tell us what model to use based on the shape of the data?

[02:40:30]
Speaker 4: You're very unlucky to ever be able to pre-specify a specific model. In the UK, the National Institute for Health and Care Excellence has a decision support unit, which has a series of technical support documents on issues that are difficult. One of them is on survival analysis, and I was one of the authors of that.

[02:41:00] It absolutely probably needs updating, to be honest, but it and other people have suggested algorithms or systematic approaches to trying to identify these appropriate models. I think you can pre-specify what types of steps you'll go through to try to identify which model you think is most appropriate. I think that's probably-

Male: A bit like pre-specifying multi-variate analysis or something like that ...

Speaker 4: Yeah, and accompanied with lots of other secondary analyses, which test whether you think your model provides a good fit. What things will you do to see whether
[02:41:30] the model provides a good fit, both to the data and beyond the data.

Speaker 6: I want to make a comment on Dr. Latimer's presentation. My understanding that his focus is to estimation. Try to get estimated average, the clinical survival benefit.
[02:42:00] You can use model to fit this whatever observed survival curve. For us, we also need to test, so right now, we still use the [inaudible 02:42:13] to make the decision whether or not this agent has an effect, significant effect. Later, once you demonstrate the significant effect, then we can talk about whether or not you
[02:42:30] should use the hazard ratio or median difference, or even we can talk about the restriction means survival time.

Male: I assume that some point down the line, you'd be perhaps specifying an algorithm to pick a model and then some tests to go with it. That's within the regulatory

possibility, right?

Speaker 6: It's possible, but right now, we're still ... We have to separate the test and estimation.

[02:43:00]
Speaker 4: I think 1 important distinction to make is that, for economic evaluation, you need a point estimate of the effectiveness of a drug. It has confidence intervals, and the uncertainty goes into the analysis as well, but you have to have that estimate. Whereas for regulatory, often it's more a question of, "Is there an effect?" You don't necessarily need to know the exact point-

Male: Those could be combined at some point.

Male: Right.

Renzo Canetta: Number 2, please.

[02:43:30]
Male: I'm [Ed Roy 02:43:29] from BMS. I'd actually like to follow up on the question and comment, actually by Dr. Ratain, and in the context of estimating survival from early clinical data, so not for a program that's early on in clinical development for which we don't have maybe large numbers and some follow-up but many subjects with shorter follow-up.

[02:44:00] The question is, "What can we do to go beyond what's being used to predict survival before we get tumor shrinkage," which was mentioned. That's only 1 feature of the tumor response. Should we be looking at more than 1 feature of the tumor response? Tumor shrinkage and, let's say, rate of growth and not at a particular time point, but ought we be looking at those features for the entire [inaudible 02:44:25] longitudinal time profile that we have, for the data that we have available?

[02:44:30] If you have data from 1 subject that's only let's say 2 months of data, you take that into account. For another subject, if you have 6 months' worth of follow-up, you take that into account as well. In the totality of the database that you have ... Is that something that we ought to be looking at?

Speaker 2: You can choose the short answer. Yes.

Speaker 4:
[02:45:00] Yeah, I think both suggestions that we've had are really important. One of the key criticisms of the type of modeling that I often do and we do in the UK is can you show the biological link between what actually happens to the tumor and what you're saying is going to happen to survival? If there were ways of modeling the underlying biology, which you can link to your estimates of survival, then that would be brilliant. All of these things that are being suggested are very good.

Speaker 6: [02:45:30] I think right for the immunotherapy, we also talk about the immune response before the clinical response. We also to have good job to how to detect the immuno response before the ... Combine the immune response and the clinical response, even the long-term effect, then we can have a clear picture to describe that is treatment.

Renzo Canetta: Number 3.

Male: I think that was a ... Oh, I'm sorry.

Female: [02:46:00] Can I just make a comment? I think that is a very good point. I'm not very familiar, or I can't think right away about how you can do it in a phase II or something, but it's very similar to a phase-I trial, where you have multiple endpoints and multiple timescales, and you are adapting your design as the data comes in to figure out, "Where do I treat my next cohort of patients?"

[02:46:30] It think it is definitely something you have to consider, right? It's tumor shrinkage, rate of growth. There is multiple components to this endpoint, and how do I use all of this as I design the trial but also making [inaudible 02:46:29] for the trial, to continue to trial or not. I don't know how to do it, but I think we should do it. I think that's maybe a very good point.

Male: Ideally, you'd actually also want to look at the pre-baseline data, to look at the change as well.

Renzo Canetta: Number 3.

Male: [02:47:00] [inaudible 02:47:24], GlaxoSmithKline. My question is about the independent review of the imaging scans. It applies, obviously, both to RECIST and IR response criteria. I wonder if the panel could comment on how for immunotherapies the role of that kind of independent review is different from [inaudible 02:47:09] chemotherapies or targeted therapies and, particularly, if the FDA's willingness to use and accept audit-based approach is where we look at just a sample rather than the whole trial population has changed.

Speaker 2: Sounds like a question for Dr. [Tang 02:47:27].

[02:47:30]

Shenghui Tang: [02:48:00] Well, for the audit analysis plan, we have did research [both 02:47:40] [pharma 02:47:40] and the FDA have published paper, and also NCI have published paper. All those are based on the traditional RECIST criteria. We don't have a complete data for the IR response, so I don't know whether or not we can apply the same results to these new criteria.

Male: That you can ...

Mandrekar: As Shenghui pointed out, we don't have experience with immuno-oncology

[02:48:30] products. I would ask, actually, Dr. [Schwartz 02:48:30] whether he thinks that the 2 should be weighed differently with respect to investigator with this independent review. What we showed was, yes there are discrepancies between the 2 of them, but when we are looking at the relative treatment effect of investigators and then independent review, the inference is going to be the same.

Schwartz: [02:49:00] I think that's generally people's experience. I don't quite even know what the independence actually is. I will say that there are some differences, be it on either side, so it's expert rather than independent interpretation, and that may be, quite frankly, more at sites, especially big sites that are seeing a lot of these cases.

[02:49:30] I think in both arms that potentially decreases the noise and increases the potential accuracy of the system. I mean, there is one difference, though, rather than independent versus non-independent, a retrospective versus a real-time assessment. In the retrospective assessment, you're able to check and look at all the time points with blinded to what treatment arm they're on and potentially make decisions that way, which, quite frankly, may be more informed and eliminate potentially informative censoring than you would in real-time.

[02:50:00] I think that, by far, those are the bigger differences than whether it's independent or site. I think it's a matter of expertise, and I think it's a matter of testing the drug-

Section 17 of 45 [02:40:00 - 02:50:04]

Section 18 of 45 [02:50:00 - 03:00:04] *(NOTE: speaker names may be different in each section)*

Schwartz: The expertise and I think it's a matter of testing the drug, not the person who's actually just happens to be looking at the images that's important.

Speaker 2: [02:50:30] I add to that, the uniformity, I think, of applying the criteria that are actually meant to be used is easier to ascertain if you have so called, independent review, call it uniform review if you want. That's really what it's supposed to serve and when I look at the stress from the subjective perspective data I have, I would say there's not much of a difference. How you apply this kind of uniform review to either resist or IR resist, I resist, I am resist or whatever you want to call it. So the complexity goes up a little bit and with that you might actually benefit from even more uniformity in the application of the criteria.

[02:51:00]

Female: Can I just make a comment? The issue with that though, with the retrospect of assessment, Larry, is you have to literally just deal with whatever data has been collected. How do you actually then eventually use that to make informed decisions and outcomes? I mean, I think I always struggle with that when I get the retrospect a review, much after the fact.

Schwartz: That's definitely a challenge with it.

[02:51:30]

Renzo Canetta: Just as a comment pertaining not to efficacy, but to safety, and not to who is

reviewing, but to what type of data is going to be audited and reviewed. There's been work done together with the FDA and with several companies about sample auditing. And this has been naturally very fruitful. It has resulted in some guidelines beneficial to submission of dossier for additional indications. So again, there is
[02:52:00] openness from everybody to discuss these types of thing and do it in the most efficient way. Number two.

Rob Ross: Rob Ross, Bluebird Buyout. So, it's a great panel, thanks guys. Actually you had a point earlier that I'm interested in a little amplification of. As we move from combination therapies where IO therapies for those patients prior to coming on weren't available, to now, a wave of trials looking at new IO treatments in patients
[02:52:30] who have failed previous IO treatments. I think your point around delayed progression or pseudo progression leading to a response, and how that affects how you evaluate the next treatment is a really important one, and perhaps a really confounding one. Especially when you have small numbers of patients. Do you or does anyone on the panel have practical strategies for how these early trials should be designed looking at patients post say, PD1 inhibition, to make sure that you
[02:53:00] aren't falsely lead down a false path based on a delayed response to an earlier treatment.

Speaker 2: Yeah, so I don't have a crystal ball, I don't know the real answer, but I can make a projection by saying you will have to deal with the real world scenario. The real world scenario is that patients get these drugs. PD1 will be given to many, many cancer patients and it will have an influence not just on the immune response, the [inaudible 02:53:26] response, it will also prime the immune system of the patient,
[02:53:30] be that a [inaudible 02:53:31] in the tumor to look different if that PD1 treatment was successful or if it wasn't. The immune system may still look different after the exposure to that agent. Now just give you an example of something that's quite obvious and many of us here would appreciate, is if you have T cells that get exposed to PD1, you up-regulate other surface receptors. And those other surface receptors can be targets for the next intervention. So if that is x40, which is actually
[02:54:00] a true example, we have x40 antibodies in the clinic, and then you can actually start beginning to look for which sub-population of patients are potentially candidates for getting this intervention versus another based on biology that we understand. So I don't view this as a disadvantage, it's just something that needs to be carefully investigated.

Speaker 3: I actually would like to point out to that fact that it's a limited experience but in the
[02:54:30] first combination trial done by the memorial and the Yale group with an [inaudible 02:54:36] there were some data concerning patients that had been previously exposed that we [inaudible 02:54:42]. And here you are both the immunological parameter, but also the pharmacological parameter, we learn yesterday that these antibodies stay in circulation for a long time and even if we move on to other fields
[02:55:00] such as cell therapy, these cells will be around for a long time. Number three is going to be the last question.

Female: Thank you, I would like to thank the panel for all of the presentations. I am an

oncologist at MD Anderson [inaudible 02:55:11] I treat patients on Phase one trials with targeted therapy and immunotherapy. The challenges we face as clinicians and on behalf of patients is continuing treatment knowing that only five to ten % of patients who are thought to have [inaudible 02:55:29] progression, in other words,
[02:55:30] of patients who immunotherapy trials have progression of only five to ten % are reported to have [inaudible 02:55:37] progression, how are we going to deal with this offering a patient treatment beyond progression when the probability is only five to ten percent? And are there any efforts to analyze the data, the existing data that I assume there should be many patients so far, to detect or identify risk factors
[02:56:00] or parameters that will be associated with [inaudible 02:56:03] progression so that we can allow continuing treating patients, only those who are expected to have [inaudible 02:56:10] progression.

The other question I have is, as we treat patients also with targeted therapies and with our experience with cytotoxic therapy, there have been many cases where patients have, in the beginning, disease progression and then their disease
[02:56:30] regresses. And we have anecdotal cases with targeted therapy due to inflammation or other reasons. The challenge we have is how we compare these trials - the trials with immunotherapy versus the trials with targeted therapy. Currently, all [inaudible 02:56:41] with immunotherapy allow patients to be treated beyond progression. So how are we going to interpret the data coming from immunotherapy and perhaps historically compare with trials with targeted therapy? Because if the same criteria applies to trials with immunotherapy perhaps
[02:57:00] there should be some flexibility to continue patients who were treated with targeted therapy at or cytotoxic therapy and that they have progression due to inflammation.

Speaker 2: Yeah so in principle, I would agree with that. There is a bit of judgment call to be made by the treating physician, what's good for the patient, because she may not want to continue patient beyond progression with a drug that has a mechanism
[02:57:30] from which you don't expect additional benefit. If there is nothing more coming, why would you want to continue? Based on what we know today, there's this period where we are switching from conventional therapies to IO agents, and there will be control groups and randomized trials where you have to compare IO versus non-IO. The time will come when it's going to be all IO. And then that problem is
[02:58:00] gone. And you're seeing it already, right, the first chemo therapies have been eliminated. The [inaudible 02:58:02] melanoma is I think, no longer a therapy that people would want to use. Practotaxin lung cancer is the same thing.

So now we're switching to the combos. The chemo combos with immunotherapies are coming. And yes, initially, there are control groups that are chemo versus chemo plus IO. But just give us a little time because what's next is chemo plus IO
[02:58:30] becomes standard of care. If you want to improve on that, your control group will be containing some IO agent and with time, I predict chemo will be disappearing. So I agree with you in principle, there's a problem here that needs to be managed, but with time, that problem will become much smaller.

Female: Are there any efforts to identify risk factors associated with [inaudible 02:58:49] progression?

Speaker 3: [02:59:00] I think all the effort that is going on is not actually solely related to [inaudible 02:59:00] progression is to do with prognostic factors for the overall outcome, that encompasses also so the progressions. Do you have a comment on this specific point?

Speaker 1: [02:59:30] I wanted to clarify, number one, in terms of progression before, if you really look at the scans carefully for patients who progress before they regress, usually you see some evidence of mixed response. It's very, very, very unusual where every lesion is progressing and then they go on and develop a late response. But there are patients that have an 80% progression and one or two lesions are starting to shrink. If you just wait four to six weeks you may start seeing regression at that point in time. It does happen. The other point is that for other kinds of agents we treat beyond progression all the time. You just saw an example for an enegiophar receptor inhibitor, we do it with the [inaudible 02:59:55] and we do it with the vegophreceptor inhibitors so that the concept of treating beyond progression goes [03:00:00] way beyond immunotherapy, we're already doing a standard of care with many

Section 18 of 45 [02:50:00 - 03:00:04]

Section 19 of 45 [03:00:00 - 03:10:04] *(NOTE: speaker names may be different in each section)*

Speaker 1: ... goes away beyond immunotherapy. We are already doing at a standard of care with many other types of agents.

Speaker 2: However the protocols do not always address that. In other words, the protocols with immunotherapy allow treatment beyond progression. However, other protocols with non-immunotherapeutic agents do not allow treatment beyond progression.

Mark: We do in melanoma. The targeted therapies allow treatment beyond that.

Mark: Talk to your IRB.

Speaker 2: [03:00:30] I'm talking to [crosstalk 03:00:25] but the question is there any interest in identifying risk factors? I'm sorry for insisting on that but ...

Mark: [03:01:00] The answer is yes and it is the short answer and people are working on that, but again not all end up through the progression and everything goes well. I like to conclude this panel with a comment if I may because we've spoken a lot about methodologies and so on. One thing that we should never forget and that applies to immuno-oncology, as it does apply to every other sort of cancer therapy. We are treating patients. Sometimes you need to re-conciliate the criteria that we will utilize for regulatory purposes also with the interest of the patient, is that the patient is doing well as Dr. Sznol alluded before.

We also should always keep in mind that we are trying to do the best interest of

our patients. Having said that, I'd like to thank the panel for a terrific set of
[03:01:30] presentation. We will resume at 12:15 after lunch.

Marc Theoret: Let's go over one thing which I think we kind of drove by at the end of yesterday's session. It was one minute we will have to talk about DLT observation periods. I just
[03:02:00] wanted to say that one of the panelists, he made a point that the DLT observation periods are often based on cytotoxic chemotherapies as opposed to actual mechanism of action of the drugs. I think majority of the panels were kind of discussing checkpoint inhibitors because that's what we have the most data for at this point at least with CDER and the IO products.

[03:02:30] It is definitely a very challenging definition to make with this current class of IO products. I think at least with the checkpoint inhibitors what we have seen is really a wide variability. There have been extended observation periods for DLT, which have been proposed really to account for the PK and potential PD interactions with the later onset of the immune-mediated toxicities, but it really is a practicality issue.

[03:03:00] One thing that we have been advising sponsors to certainly do is to account for the later onset of immune-based toxicities that are dose limiting in the recommended phase II dose in description thereof, as well as depending on the level and severity of toxicity, potential stopping rules within the protocol to handle these issues especially with combinations. I think Dr. Feltquate had gone through that very well with the IPI and NIVO story and what they were doing in terms of trying to
[03:03:30] determine the appropriate dose to move forward into larger numbers of patients.

I'll say fortunately at least with this class of the IO products, anti-PD-1, anti-PDL1, at least with the information that we have. The exposure response analysis for safety have been relatively flat and toxicities have been very well managed with immunosuppression. That has been helpful in that overall assessment for
[03:04:00] determining an appropriate DLT observation period. In terms of this session, I just want to start with a few comments. We are going to keep the introductions brief because there are actually seven presentations to go through before we hit our panel session.

Just for time considerations, we will keep those brief. This will be the discussion of novel endpoints. Use of alternative FXC endpoints. Just I want to make a few
[03:04:30] comments before we start this session regarding endpoints and we heard very nicely from doctors, Hazarika and Mandrekar this morning regarding regulatory framework for approvals as well as regulatory considerations with the current class of IO products at least from CDER regulated products, checkpoint inhibitors.

I do want to just stress that tumor response base endpoints really do permit earlier valuation of anti-tumor activity. With objective response rates and durations of
[03:05:00] response really permitting evaluation of direct drug effects on those tumors and really one that has frequently supported accelerated approvals. Just looking across the new molecular entities that were approved through the offices of

hematology/oncology products 2014-2015. There were 24 new micro entities for the treatment of oncologic diseases.

[03:05:30] Out of this 24, 12 were accelerated approvals. Of those 12, that accelerated approvals, 9 out of 12 were based upon an objective response rate and duration of response. Very important, even as an earlier intermediate clinical endpoint or earlier endpoint I should say to support other types of expedited programs such as breakthrough therapy designation, objective response rate is very critical. Of those
[03:06:00] products in OHOP that were granted breakthrough therapy designation up through 2014, which I think is the last time was probably presented it was about two-thirds of the breakthrough therapy designations were supported based upon objective response rates and duration of response.

Really a critical new drug development tool and one for which we have made regulatory decisions based upon of. It is a standardized response criteria that does
[03:06:30] facilitate that evaluation of a new drug activity versus historical controls to permit these single arm analyses and potentially expediting drug development from that perspective, but there is certainly a great interest for us in really the use of efficacy endpoints that can provide earlier evidence of clinical activity.

What I've always go back to for myself and what have shown by multiple speakers
[03:07:00] is really the ipilimumab story, where with very low objective response rates, 6% versus control and progression free survivals, which at least at the median were very close to control. We really would may not have had this therapy, had it not been for the trial evaluating OS and demonstrating a significant impact on that endpoint. To proceed, we are going to go ahead with the sessions in the panel, the
[03:07:30] presenters and then the panel. I would like to invite the first speaker, who will be speaking about modified progression free survival as a potential surrogate for survival immunotherapy, Dr. Srisha Mushti.

Srisha Mushti: [03:08:00] Thank you Marc and good afternoon everyone. As Marc just pointed out, it is a very important thing to know which of these endpoints would be like good estimates of treatment benefit or not. As an expression towards those, we will be looking into two different aspects of modifying the PFS and a novel response endpoint. We will be discussing about approach towards the exploration of novel endpoints. That could be used for evaluating immunotherapeutic agents. Now this consists of two
[03:08:30] parts. The first on the investigation of modified PFS.

This is a joint work by Ms. Mulki, Dr. Sridhara, and myself and the second on a novel response endpoint and this is a joint work by doctors Ghou Zhang and Dr. Sridhara from the division of Oncology at CDER FDA. As discussed in the previous presentation since this morning, it was evident that there were concerns with RECIST-defined FPS, and there is definitely a need to explore new or intermediate
[03:09:00] endpoints. Now in this part of the presentation, we will be focusing on the modification of the RECIST-criteria to redefine the PFS endpoint.

Here is the triple disclaimer. The views and opinions expressed here are those of

the authors and should not be considered as those of FDA's user policies. Moving to the first part, we know that the traditional endpoint of PFS is derived based on
[03:09:30] the RECIST defined disease progression. Now by RECIST criteria, at least 20% increase in the target lesion tumors from [NADR 03:09:37]. Development of a new lesion or a progression in non-target lesions are considered as disease progression. Now whenever one or more of these criteria are met, typically the current treatment maybe discontinued, and the patients can receive their alternate treatments later on.

[03:10:00] However, when patients were treated with immunotherapeutic agents such anti-PD1 or PDL1 agents ...

Section 19 of 45 [03:00:00 - 03:10:04]

Section 20 of 45 [03:10:00 - 03:20:04] *(NOTE: speaker names may be different in each section)*

Srisha Mushti: Agents, such as Anti-PD1 or PD-L1 agents, a distinct immune-related pattern of responses were observed, such as initial increase in the tumor lesions, later followed by subsequent decrease in the total tumor burden. Now, these are the [inaudible 03:10:16] spots of pseudo-progression, which are being classified incorrectly as progression at a very early stage during the treatment period. Also,
[03:10:30] delayed clinical responses were observed later in some patients treated with immuno-oncology drugs.

So, we attempted to address this problem of pseudo-progression by cancelling the extent of change in the tumor size of target lesions. So, as one possible solution, instead of looking at 20% increase in target tumors, we explored different threshold values, beginning from 20%, 30, 35, 40, and 45%. And the definition of
[03:11:00] PFS is still the same, as defined. So, we know that it's defined as time since progression to disease progression, or death, by using the new cutoff values to define the disease progression.

Now, once we defined these modified PFS, it is important to evaluate whether these endpoints can be considered as a surrogate for the long-term treatment benefit for the long-term survival benefit or not. So, we investigated the surrogacy
[03:11:30] criteria for these modified endpoints, in relation to OS, using the trial level and the individual level of associations. And this was conducted as a meta-analysis, using the data from the clinical trials of Anti-PD1 and PD-L1 agents, which were submitted to FDA as either initial or supplemental BLA, across multiple indications since the year 2014.

Now, the eligibility criteria for the trials to be included in this meta-analysis were
[03:12:00] they need to be randomized, multicenter, or active control trials. Now, from our database, we have identified 11 trials, which included approximately 6300 patients, and consisted of 15 comparisons obtained by treating the three [inaudible 03:12:17] trials as two separate trials. Now, the active control trials could be designed either as a head-to-head comparison, or as an add-on to the standard of care background.

[03:12:30] So, we have evaluated the surrogacy criteria of individual level of association, using the rank correlations between the modified PFS values and the sub-value for each patient, and at the trial level associations. So, we have cancelled the associations between the treatment effect of PFS and OS ... using the [inaudible 03:12:57] linear
[03:13:00] reduction model here, with the weights equal to the number of patients. So, the R-squared here ... tells us what is the level of association between these endpoints.

Moving into the data, these are the trials that we have considered. These are the 15 different studies that I was just mentioning previously, and we have even listed the primary and the secondary endpoints for each of these studies, along with the number of patients involved in each comparison, the hazard ratios, the unstratified
[03:13:30] hazard ratios for the overall survival and PFS. Now, this PFS is as defined by the RECIST criteria.

Now, to understand the impact of the modifications that we have considered in our work, let's look into the number of PFS events, as defined by the RECIST, using 20% and the modified PFS cutoff values. Here, I would like to draw your attention to the
[03:14:00] first row here, which includes the total number of PFS events from the pooled studies. Now as you can observe, there's definitely a decrease in the number of events as the cutoff values increase. And this is because the modifications that we have considered is mainly impacting the number of disease progressions. And we know that ... And also the other point I would like to mention here is the difference in the percentages are not too big and that's because the patients, who were
[03:14:30] classified to have progressed based on target tumor, the majority of them progressed because sufficiently large increase in target tumor size, which would be even more than 45% also.

So, how many such patients were there, who were contributing to these differences? Here's a quick insight into that. We note the progression is the PFS events as defined as either death or PD, due to target, non-target, or new lesions.
[03:15:00] Now, the differences that we saw earlier were due to only the target tumors. For example, if you are using the 20% cutoff here, almost 51% of the patients were defined to have PD, due to an increase in the target tumor size. Now, when we change the cutoff value from 20% to 25%, some of the target lesions, whose increase in the tumor size was between 20 and 25% are being excluded, when we
[03:15:30] are computing the number of events using the higher cutoff values, like 25% or 30%.

Now, a little more details about how the percentage changes from baseline and target tumors were during the first assessment period. Now, the blue bars here corresponds to the proportion of patients in control arm, and the pink for the experimental arm. The right side of this 20% increase reference line, such as
[03:16:00] increase in tumor sizes and there doesn't seem to be much differences between the arms in tumor response, except for few during the first assessment period. Now the positive responses, which are greater than 30% decrease, are shown to the left of this 30% decrease reference line. Now, more patients were showing positive responses in the treated group. Now, deeper positive responses were

observed in the experimental arm than the control arm here.

[03:16:30] So, with these insights, we looked into the trial results. For the individual level associations, as mentioned earlier, the rank correlations between each of the modified PFS and OS, were calculated. And you can notice that all these numbers are almost close enough to each other, implying that the cutoff value had really no
[03:17:00] big impact on the association values, between the PFS and OS. Now, sorry, just going back ... One more point here is also the values are closer, like 0.5, so the association at the individual level were not very strong.

Now, moving to the trial level, we first looked into the association for the traditional PFS endpoint, as defined by RECIST, this was also presented earlier by
[03:17:30] Dr. [Shedrine 03:17:25], her presentation. And in this, we have found that the R-squared values were pretty low, .078. And even the log-transformed value didn't turn out to be too good, with respect to the R-squared values.

So, moving to our modified PFS values here, this line shows the trial level association between each of the modified PFS and OS. The R-squared values reported here are based on the log-transformed of the hazard ratios. The
[03:18:00] correlation values are very low here, ranging between .13 to .18, for each of the different cutoff value considered. So, really like using a more stringent criteria for PFS, did not result in a better association of PFS to OS here.

So to conclude, we looked into the tumor changes, like the changes in tumor burden, based on target tumors to modify the PFS endpoint. The surrogacy values
[03:18:30] resulted in weak associations between the modified PFS and OS, as was the case observed with the RECIST-defined PFS. Now, some of the limitations of this work were differences in the follow up past progression was different in the control arm and the treatment arm, but to some extent, it was not too much of a difference because it was only few protocols where the patients were not followed past progression, but for the others. And definitely, all the treatment on patients were
[03:19:00] followed past progression in here. And the other caveat that we had is the evaluations were only based on positive trials, and no negative trials were included because it was not basically available to FDA. In addition to looking into the modifications, with respect to the PFS endpoint, we have also looked into the novel endpoint based on tumor responses and this will be presented by our next speaker,
[03:19:30] Dr. [Gao 03:19:25]. Thank you.

Xin Gao: So, good afternoon, everyone. I will present some analyses we performed to explore a novel response important for immunotherapy clinical studies. So, the work is jointly done with Dr. [Lee Jin Jeong 03:19:49] and Dr. Rajeshwari Sridhara at FDA.

[03:20:00] Okay ... alright, this way.

Section 20 of 45 [03:10:00 - 03:20:04]
Section 21 of 45 [03:20:00 - 03:30:04] *(NOTE: speaker names may be different in each section)*

Xin Gao: All right. This way. As other presenter discussed, although OS is still the gold standard for cancer trials, it may not be feasible to be used as a primary endpoint in future clinical studies. With the concerns of the RECIST criteria, such as the response after progression, pseudo-progression, and the response in the presence
[03:20:30] of a new lesion. PFS and ORR may not be the optimal primary endpoint either.

Therefore, there's an urgent need to develop an alternative intermediate endpoint, to be used in immunotherapy clinical studies. In this research work, we aim to develop a novel endpoint, which could predict the long-term survival benefit. Similar to the RECIST-based response rate, this intermediate endpoint will still be
[03:21:00] based on the radiology image data. Beyond that, this endpoint will also incorporate information which is missing from the RECIST based response rate. Such as the information about tumor change dynamics, including the duration of tumor shrinkage.

To develop and assess this intermediate endpoint, we first collected individual level data from multiple anti-PD-1 immunotherapy studies. Then developed this candidate intermediate endpoint, using a subset of the collected data, which will
[03:21:30] be referred as the training data-set in the rest of my presentation. The evaluation of this candidate intermediate endpoint was performed through an individual and trial analysis, using all the data that we collected. The individual analysis evaluates the association between the 2 endpoints, the intermediate endpoint and overall survival. The trial level analysis evaluates the association between the treatment effect on the candidate intermediate endpoint and treatment effect on the overall
[03:22:00] survival.

The data we collected included 9 anti-PD-1 immunotherapy trials. This 9 trials have overall survival as either the sole primary endpoint or as one of the co-primary endpoints. The indications of this 9 trials, cover 4 solid tumor types. The trials are well-sized, with the largest one having more than a thousand patients. 4 studies
[03:22:30] were randomly chosen to be included into the in-training data-set. The remaining studies were used as independent testing data-set to assess intermediate endpoint.

This intermediate endpoint was developed as a composite endpoint, consisting 3 factors: target lesion, non-target lesion, and new lesion. Among this 3 factors, for the target lesion, is summarized with 3 sub-factors: baseline tumor burden, post-
[03:23:00] baseline reduction DAFs, and post-baseline tumor change dynamics. This 3 sub-factors were used to determine the target lesion response. I will give more details in next 2 slides.

Baseline tumor burden was measured by some of the longest diameters of a target lesion, at baseline. Post-baseline reduction DAFs with target lesion later, within the year on a scale of percentage change from baseline. For target lesions, post-
[03:23:30] baseline lesion change dynamics was represented using the area under the curve over time from baseline, within a year.

This figure illustrate an example, how to calculate a patient area under the curve.

[03:24:00] The x-axis is the time in days, and the y-axis is the percentage of change from baseline. The circles and the dots on the curve, represents the repeated measurements of the percentage change from baseline, for each subset, at each assessment in the trial. Then the smoothed curve can be obtained, and area under the curve is determined by adding the areas below and above the baseline. The area below the baseline is treated as a negative value, reflection a reduction of the post-baseline tumor burden. Area above the baseline is treated as a positive value, reflecting an increase of the post-baseline tumor burden.

[03:24:30] For example, in this figure, the area under the baseline is represented as the area 1 equals to -3623. The area above the baseline is represented as the area 2 equals to 2018. In this case, the area under the curve is a sum of area 1 and area 2, which equals to 1605.

[03:25:00] We've done fitted a cause regression model on all our survival and include, this 3 sub-factors into the model, using a training data-set, and determined the coefficient for each factor. We evaluated the predictability of this survival using C-index. A value closer to one, indicate a better predictability of this value. The C-index was 0.79, using the data from the training data-set. Then we used the same functional form of the model with the coefficient to obtain a target lesion score for

[03:25:30] each patient, and then explore within training that's different target lesion cut-offs, to identify an optimal cut-off that yielded the maximum difference in the OS curves, in order to define a target lesion response criteria.

[03:26:00] Based on the data, this optimal cut-off was determined to be one. The overall intermediate endpoint is defined as a binary, response or no response. Here, we defined response as satisfying all 3 criteria,

First, the patient needs to be a target lesion responder. Meaning the patient target lesion score is less than the optimal cut-off, that we just identified. The patient has no progression of non-target lesion within the year. Third, the patient has no new lesion within the year. With this definition, we determined the intermediate response data for every patient in the analysis data-set.

[03:26:30] We then, assessed the association of this candidate intermediate endpoint with OS, through an individual and trial level analysis. Let me show the individual level analysis first, which evaluates the association between these 2 endpoint, intermediate endpoint and overall survival. This individual level assessment answer the question, whether a patient with intermediate endpoint response will have a longer survival outcome.

[03:27:00] This took couple of Meier curves of all the survival represented by the intermediate endpoint response data. The left figure shows the results using the training data-set. Couple of Meier curves separates from the beginning. The results were confirmed using the testing data set shown in the right figure. This implies, at individual level, the candidate intermediate endpoint was indeed, associated with

[03:27:30] the overall survival. Individual patient who attains an intermediate endpoint

response have a longer survival outcome.

[03:28:00] Please note that this comparisons are not based on the randomized group comparison, which are necessary for any statistical inference. The trial level analysis evaluates a more important question, the association between the treatment effect on the intermediate endpoint and the treatment effect on the overall survival, and answer the question, whether the magnitude of treatment effect on the intermediate endpoint can predict the OS benefit. We examined association using a weighted linear regression model, with the sample size as the weight.

[03:28:30] In this plot, each circle stands for one randomized comparison. The odds ratio of intermediate endpoint, and hazard ratio of overall survival was estimated from each randomized comparison. Different color indicates different tumor types. The size of the circle is proportionate to the sample size of the trial. The coefficient of determination, R-square, was calculated to marry the correlation between the effect size of overall survival versus intermediate endpoint. R-square closer to one, indicates a stronger association. This plots indicate that, the larger odds ratio of the intermediate endpoint is, the smaller the hazard ratio of overall survival is.

[03:29:00] Using a weighted linear regression model this R-square was estimate to be 0.65. Indicates that at the trial level, there's a moderate association between the treatment effect on the intermediate endpoint and the treatment effect on the overall survival. Please note that as presented by Dr. Sridhara this morning, the r-square was 0.1 when the traditional response rate was used.

[03:29:30] In summary, we developed an image database intermediate endpoint, which consider 3 factors including the patient target lesion, non-target lesion, and new lesion. The target lesion response is summarized by using the baseline study of tumor, post-baseline reduction DAFs, and post-baseline tumor change dynamics. The preliminary results from the individual and trial level analyses indicate that, individual level patient who attains an intermediate endpoint response has a more favorable survival outcome. At the trial level, the preliminary result shows a [03:30:00] moderate association between the treatment effect on the developed intermediate endpoint and treatment effect on the overall survival-

Section 21 of 45 [03:20:00 - 03:30:04]

Section 22 of 45 [03:30:00 - 03:40:04] *(NOTE: speaker names may be different in each section)*

Xin Gao: [03:30:30] And the treatment effect on the develop intermediate appoint, and treatment effect on overall survival. Although this is not a perfect result yet, but the result is better than the PFS, and ORR results using the traditional resistor criteria. This is still ongoing research, and the future work remains, such as refining the [target-leader 03:30:19] model, or evaluating the different cutoff for the imaging data. That's it.

Marc Theoret: Thank you for that, those two presentations. I'd like to invite Dr. Toni Ribas up to speak about how we assess benefit from immuno-oncology agents.

Antoni Ribas: [03:31:00] The first thing ... Congratulate Dr. Hazarika, Theoret, Topalian, and Wolchok, for putting together this meeting and inviting me to present in it. The question they posed to me is how do we assess responses to immuno-oncology agents? When they posed this question, I thought there's one thing we're trying to achieve. The reasons we're all here is because we know that immuno-oncology agents can lead to a response, and that response can lead to a long-lasting response. The immune system remembers, has memory, and that's why many of us in this room have been [03:31:30] sticking with this idea and working on it, even when it wasn't [ascribed 03:31:34] now.

Our goal is to get a patient to respond, and maintain the response long-term. I'm going to try to capture this idea with one anecdote, but first I'll go through my disclosures. Obviously, I don't work at the FDA, because I have a whole bunch of conflicts.

[03:32:00] I'll take you back ten years. I don't usually read Forbes, but I read this issue, because it talked about the magic bullet for melanoma. It was reporting on some patients that [Detvolchok 03:32:08] and I were treating. This is from before the economic crash, so I wish I read this other one that said the recession of 2008 ... This guy's got predicted, so I wish I had read that one, but I read this one.

[03:32:30] This is a patient of mine, who, in 2004, was 39 years old, a physician from New Zealand, who had received the treatments there. Had metastatic melanoma to the liver, adrenal glands to the bone. Something that wasn't curable at that time. And the patient received tremelimuamab, could have received tremelimuamab. It doesn't really matter. The CTLA-4 blocking antibody.

When the patient first came to see me he did not say "Hi Tony. I would like you to [03:33:00] improve my hazard ratio for .68." He understood what hazard ratio meant. He understood clinical trials, but that's not what he's asking for. He's saying "Tony, I want to see my kids grow up. I want to see my two girls go to college, and help them to go there." Obviously, I am showing you this because I did do that. I was there a year ago with active volcano called Rangitoto. Shortly after, he was going to take his oldest daughter, who was now in college time, to college in the South [03:33:30] Island of New Zealand. This is what we should achieve.

We should not decrease our level of expectations. We can design and point on clinical trials that capture something else. But these are the ones we want to capture.

Objective response rate does capture it, but we talked long that it has shortcomings. PFS, I don't know if it does or not because if we cannot define objective response, then how long the progression and we'll have troubles defining [03:34:00] progression. Overall survival, has a whole bunch of issues. Mostly, post progression therapies that may alter that, now we have multiple active therapies. Once concept that trial was brought up a while ago was durable response rate, which Dr. Hazarika

also brought up early this morning. Which is you have a patient with at metastatic cancer. Given an I-O treatment. the patient starts responding. Responding and
[03:34:30] responds long-term. If we can capture this and an endpoint, then we'll be really capturing what we are trying to achieve. It doesn't really matter if there is a time period the lesions look bigger or the new one appears. It doesn't really matter, as long as without changing the treatment, the patient gets to this level.

[03:35:00] There was a regulatory ... Maybe other trials and other regulatory agents know that or colleagues from the FDA ... But this one which is the [T-Beck 03:35:07] trial vs. GMCSF use durable response rate as it's endpoint. It was a durable response rate that was low, but still, was positive the control arm at the very low frequency of durable response. [T-Beck 03:35:22] had the higher. There was a selection of the population that had a whole bunch of issues. But probably this agent would not be
[03:35:30] approved with anything else. I think this endpoint does capture what we are trying to do because the more traditional endpoints of PFS and OS work great when you have a therapy that benefits the majority of patients a little bit. As exemplified by targeted therapy. So if you select a population of patients and give them a treatment that has a benefit for a period of time, the majority of patients have
[03:36:00] some benefit ... Let's say 50% shrinkage of the tumor ... But if that's not long lasting, then you have PFS and OS but there is no durable objective responses.

We've known from the [tremlimuamab 03:36:14] curves that this is something else we want to capture, which is there is a majority of the immuno- therapies that have no effect on part of the curve. The effect is late. If you look at the median it is hard
[03:36:30] to capture. Because, the biggest benefit is in a few patients. The one's that have these long-lasting responses. So, PFS may be negative. OS eventually is positive. You have a hazard ration that then eventually you can demonstrate that the curves are separate. The durability of the responses, the patients who shrink the tumor and maintain that response, even though it's not a big number, it's clearly
[03:37:00] measurable.

Where it gets tricky, is now that we are doing randomized trials not against a therapy that is like this and this, but now, we are going to do clinical trials comparing this with this. Now we'll have curves that we discussed earlier in the session before. Which is how do you manage when you are comparing a therapy that has early benefit in the majority of the patients from a therapy that doesn't
[03:37:30] have the early benefit in the majority of patients because most of the benefit is in the minority. But those minority, get a lot of benefit. Here's where this mixed model endpoints will be of relevance. It is very interesting from the prior session to be able to learn about the statistical design from this kind of analysis.

What I'm trying to convey is that there is a whole mark for immuno-therapy, which
[03:38:00] is that you have long lasting responses. That is what we are trying to achieve. With the added benefit of one [inaudible 03:38:03] blocking antibodies and combinations, we now have ... We are increasing the power and the frequency of this event, but we still have to be able to capture it.

[03:38:30] Objective response PFS and OS are not always captures this benefit. PFS and OS usually favor a small benefit in a lot of patients, because that is the easiest to shift the curve and have an improvement in hazard ration. As opposed to having lines changing effects in some of the patients which is what we are trying to achieve.

[03:39:00] Durable response rate or maybe the intermediate endpoint that we heard from Dr. Gao, which I did not know about, does capture this intended benefit of immuno-oncology agents. It's independent of the performance of the control arm. If the control arm does the same, then great. Then we have two active long-lasting therapies. It's also independent of post-progression therapies. We have a series of active therapies that we have to worry about post-progression and the impact on the overall survival.

Thank you very much.

Marc Theoret: [03:39:30] I'd like to thank you Toni for that excellent talk. I'd like to welcome Jan Bogaerts. Dr. Bogaerts from EORTC is going to speak on milestones for survival; overall survival versus earlier endpoints.

Jan Bogaerts: [03:40:00] Good afternoon. So I'm a European as opposed to most of you. I have no conflicts, but that I should mention that the organization that I work for, EORTC, works with most major pharma companies and receives grants in various forms. I'm a permanent member of the EMA-

Section 22 of 45 [03:30:00 - 03:40:04]

Section 23 of 45 [03:40:00 - 03:50:04] *(NOTE: speaker names may be different in each section)*

Jan Bogaerts: I'm a permanent member of the EMA Scientific Advisory Group, so I guess EMA is thinking about all of this as well.

[03:40:30] For last 12 years I've been working actively on RECIST. I want to give some witness experiences from working on RECIST. Working on RECIST is a little bit hard because people start to see you as a lawmaker. The one thing that never occurred to me during this 12 years and told me, "You know Jon, RECIST is exactly correct." That's the one thing that is not going to happen. It's a constant avalanche of things that are wrong with it and that's because it is a very-, it's a bar that we've made for evaluation of results. I will talk about RECIST and irRECIST a little bit, then the lack of gold standard for these earlier endpoints. I will have some comments and then maybe some thoughts about how survival data can be interpreted. I'm already in orange color but I'm going to disregard that.

[03:41:30] The current status of RECIST, we're busy writing up the analysis of a targeted agent warehouse analysis. We had the original RECIST, which was a evolution from WHO which was really trying to set a bar, a bar for how to evaluate drugs in phase II with response rate. That evolved into RECIST 1.1 that we published in 2009. Where we really had to take into account that in the meantime, endpoints like response rate and especially PFS, had become the de facto-, for RECIST the de facto standards for phase III evaluation. That had not been the original intention.

After that, we made - a little bit late, you might say, but that's how we did it - we made a warehouse of about 15,000 patients for targeted therapy and we're now about to publish that. In a way, because there is an expectation that RECIST is evaluated on large volumes of data we're always running behind. What we are now
[03:42:30] doing is, because we don't have that much data in our hands in the RECIST group for the therapy situation, we've written recommendations for how to use it-, how to use RECIST in immunology setting. Expanding it, and trying to make a logical expansion, and that is the irRECIST that you saw in some of the tables that were shown this morning. This is being coordinated by Lesley Seymour from the Canadian NCI.

[03:43:00] I have two hopes for this publication, which will come out early December. The first is that it might give some unification, and I understand that this is another medal in a continuum of medals. I'm hoping that because it has that magical RECIST name or that magical committee behind it that this will be adhered to in some form. My
[03:43:30] bigger hope is a big part of that publication is devoted to explain what we propose as a common standard of data collection. It goes beyond what we do today. As several speakers this morning explained, when you're analyzing these RECIST warehouses you're very quickly confronted to the fact that RECIST data collection is self-limiting because it is per RECIST. As Larry also mentioned this morning, as soon
[03:44:00] as a new lesion turns up the tap is closed. Even for that date of evaluation, the data stops. No further data. We would be hard pressed to do the type of analysis that you have done, where you go out to 50% increase cause we simply can't do it. As soon as one of the RECIST progression criteria hits the data stops. Then, the next data point that we have is the overall survival endpoint, often with a significant
[03:44:30] time period in between.

Where I was asked over the past years to evaluate potential changes to RECIST, that was also within a certain window of possibility. Like simplifying it. Can we do it a bit simpler? Can we take this away? Can we tweak this or tweak that? Not going further out longer because we simply didn't have the data. My biggest expectation
[03:45:00] of this irRECIST publication is that everybody will adhere to a new model of data collection that will inform us in the future, in the coming years will inform everybody, whatever version of criteria they want to use but at least give them the opportunity to properly evaluate it.

Just a little plug. This will be discussed at EORTC-NCI-AACR molecular targets and
[03:45:30] cancer therapeutics in Belgium and Munich at the end of the year, but to us the end of the year. I think it's first of December.

I want to highlight the contrast, in my opinion, between progressive disease per any methodology and the decision to start treatment. From the patient perspective - and I'm not a medical doctor, so I'm not the patient either - it seems to me that
[03:46:00] the biggest effect to them is the decision to stop treatment. I know that many-, in many hospitals and many research groups as soon as you hit 21%, that's where the knife falls. This is the decision to stop treatment. I've never been very certain that

this should be so closely linked up. For me, progression per RECIST is an endpoint. It's something you can calculate. The decision to change the treatment of the patient is a clinical one and I've always felt that that was going one step too far. For
[03:46:30] me, RECIST is a method to standardize and reach a degree of comparability across studies.

I'm not sure it should always define the end of treatment. Indeed, what we see now as proposals is after a first signal of potential progression, proposals to go further then evaluate whether that first signal was a true progression and only then decide to stop the treatment. Of course, for all of us the big question there is, "How
[03:47:00] far can we take this?" Because while there may be some good in that decision to continue, there may also be some wrongdoing in that because as many speakers before me have said, many of these progressions are going to be true progressions. These patients are not being served, I assume, they are not being served by the decision to continue treatment.

[03:47:30] I think this is really also concentrating on giving proper guidance to say what is the right point in time to stop the treatment. I don't think that's my business. That's for the doctors in the room. I've already mentioned this, that RECIST is a paradox by construction. As soon as you reach RECIST progressive disease, the data flow stops making it impossible to look into many interesting questions just because the data is not there.

[03:48:00] Another point I want to bring out is that we have to be fairly up front about these intermediate endpoints. To an extent, they're always going to be a convention. It's like saying that .05-, less than .05 is statistically significant. It's 1 out of 20. You make of that what you want. It's just .05. It's a number like another number. There is nothing sacred about this number. We've seen that in previous presentations if you start tweaking the rules - a bit to the left, a bit to the right - not that much
[03:48:30] changes. I think the decision, the decisions that were taken on the original RECIST had more to do with the concern around precision of the imaging at that time. I've been playing with the idea to narrow it down actually because I was assuming there was more precision now. On these, that's not enough of a rationale to do that.
[03:49:00] There will always be some degree of convention in there, as opposed to overall survival. That's not a convention. That's a fact.

We can reach high degrees of correlation with overall survival, if not surrogacy in the printed sense. I wanted to also say that this is, on the targeted database, this is
[03:49:30] for all the patients together. On the vertical axis, you have degrees of response. Instead of taking the usual classification, we cut it up further to show you that there is a continuum in overall survival hazards ratio. To an extent, this kind of graphic validates the drawing of waterfall plots because it shows that there is really a continuum from complete response to lesser and lesser and lesser response, to higher and higher progressive disease. There is one outlier there at 0%. Those are
[03:50:00] patients who were not-

Section 23 of 45 [03:40:00 - 03:50:04]

Jan Bogaerts: Outlier there at 0%. Those are patients who are not completely correctly reported. That is to say there was a degree of laziness in the reporting. Someone reported always exactly the same number, which is suspicious. You see that coming out of the analysis.

[03:50:30] The point I want to make is, if I decategorize at some point and I then start moving over one category at a time to the left or to the right, those resulting curves will remain separated because of this continuum. There is no magic area where I can cut this and say here is where it really splits. I just want to show you this continuum.

[03:51:00] In this experience of being told that resistance is not a good method, or not an optimal method which it certainly is not. Sometimes the criticism sounds like, "I have another method and I have patients where I know there is activity and I can identify it. I have a higher response rate than you. So my method is better." I think that's a false argument. It's not difficult to make the response rate go up by changing the rules. We really have to target a meaningful definition for these intermediate endpoints.

[03:51:30] I was very interested in the presentation before me. It's very short to react to that. To me it seems that is going the direction of useful regulatory endpoint. But this is not something you can calculate on the spot for a patient who is in treatment because you need data further out. But at a regulatory level to have an intermediate endpoint, this kind of construction is indeed very helpful.

[03:52:00] In identifying the shortcomings of response rate, I found this table very interesting. These are patients who are alive for at least 5 years, and they are tabulated by their best overall response. You see patients with stable disease, progressive disease, unknown somehow who were alive for 5 years. This is contrasting with the analysis that was shown this morning where the non-responders who are equal for both [03:52:30] arms and the responders were very high up there. I am trying to figure out how this compares to this table, because I really felt that this table is showing me that we are missing things.

[03:53:00] While overall survival outcomes dominate, we should prepare encounter better goal setting, what is an acceptable delay or waiting time for patients. I already said that. The definition of progression should not take a form that recognizes any level of activity. I am sure we all agree about that. This question indeed will come back when overall survival is no longer significant and we are in an area of, let's call it fine tuning.

[03:53:30] I want to show you one last fault experiment. I am trying to be a bit disruptive here. The green curve is on another axis than the red and the blue curve. The green curve is actually the hazard ratio and you see that it is diverging. Initially it is one, so nothing is different between the two treatment arms. Then it diverges to a value of a half. I did that totally out of the blue. Then you see the type of blue and red curve

where initially the two curves overlap and then they start separating. This is the type of situation we have been talking about.

When I saw these curves my first reaction, and I have to say it's still my reaction, so it's no longer only my first reaction, is that this can equally be seen as a signal that the treatment is not working immediately, as a signal that there are a number of
[03:54:00] patients for whom this trial is not doing any good. These are two ways of saying the same thing I suppose. So I did something very silly. I made a mixed model if you like, in the most rudimentary way possible. I made two overlapping arms of that prognosis here. The blue is under duress so they're equal. I make a 50% bad
[03:54:30] prognosis group with nothing going on, and then a 50% group with better prognosis and really huge effect. That is much better than what you see here. You already guess what I am now going to do. I'm going to put these two together and I get the same thing, which is a very cheap trick from a statistician. I call this a caricature.

It deserves reminding because we can come up with many ways why it would be
[03:55:00] like that. Scientifically speaking we should rule out all those ways. I can think of many reasons why I would see such a curve. There may be a late drug effect, and we know this is at work. This has been fully explained in the last 2 days. There may be an unidentified factor. A bad prognosis factor of patients where a drug does not work. It may be a too broad population, which is another way of saying exactly the same thing.

If I come up with some form of noncompliance I can make the same curve actually.
[03:55:30] There may be a short term effect that is not maintained. I'm not stating that this is what is going on in the curves that we have. I'm just saying it would be worthwhile when we think from the position of patients going into these trials, that we do enough work to be able to identify maybe who are these patients. Because if I can keep these patients out of the trial, what you will see is a curve that starts here. It
[03:56:00] keeps very high and it then starts splitting as you would hope it to split. It would start from 100% splitting at a later point in time. If I can know who these patients are. I know that's a very big question. I just want to say that that is what we should strive to do. I may be a bit lazy. I'm less concerned about the exact right way to analyze this statistically. I'd rather go find out how I can find these patients beforehand.

[03:56:30] With that I want to thank you all. I'll stop here.

Marc Theoret: Thank you Jan. That's a very interesting talk. Especially, there's certainly a lot of work in trying to identify those patients who are going to respond best to these
[03:57:00] different therapies. I'd like to welcome the next speaker, Dr. Dan Chen from Genetech Roche who will speak on non-classical response patterns, immune modified resists and immune modified PFS.

Dr. Dan Chen: Thank you so much to the meeting organizers from the FDA and AACR for inviting me to present. On behalf of my coauthors, my name is Daniel Chen. I am an

immunologist and melanoma oncologist from Stanford, now at Genetech Roche.

[03:57:30] As we have discussed over the last 2 days, the anticancer immune response is a highly regulated process. One way to describe this incredibly complex biology is through the cancer immunity cycle. This is a framework that we've use to compile all of the biologic information that we understand about the immune process as it relates to cancer.

[03:58:00] This was a nice framework to start with but we all recognize that the true immunologic biology is far more complex than that. All you need to do is talk to any patient who suffers from autoimmunity. Talk to them about how their specific biology fluctuates over time.

[03:58:30] One of the main reasons for why that is is that the human immune response responds to a very involved and complex set of systemic signals. This can be further described on a population basis. That is, not all patients are the same. Certainly not all tumors are the same and not all environmental stimuli are the same. Not only are they not the same at baseline, some of these factors can fluctuate over time. That is described by the figure here on the right. This is part of a publication that myself and Ira Melman have pending. You'll note that the many factors that can help drive whether you have an active or inactive anticancer immune response,

[03:59:00] including post genetics, age, microbiome, the presence of a viral infection. Even so much as sunlight exposure and certainly immune modifying drugs. We think of this as being a complex set of tumor, host and environmental factors that govern the strength and the timing of an anticancer immune response.

[03:59:30] It shouldn't be surprising what we've observed in the history now of active anticancer hypnotherapy. These two images come from, one from an anti-CTLA4 agent, one from an anti-PD-L1 agent and again shows the kind of dramatic changes that can be seen in patients undergoing these therapies with non-classical forms or patterns of activity or response. As Dr. Rebus nicely identified, this may not be the most common pattern of non-classical response that we see. In fact we took an

[04:00:00] approach of looking back now through our large database of patients treated with atezulizumab our anti-PD-L1 agent

Section 24 of 45 [03:50:00 - 04:00:04]

Section 25 of 45 [04:00:00 - 04:10:04] *(NOTE: speaker names may be different in each section)*

Daniel Chen: Patients treated with atezolizumab anti-PD-L1 reagent to try to better identify the many types of patterns that could be present. This slide shows four such patterns, and you'll note that while IRRC was an incredibly important attempt to start to capture these non-classical responses, that system, itself, does not capture the incredible breadth of non-classical responses that can be seen, and that can be

[04:00:30] noted just by the upper left-hand panel. You'll note that that patient experiences three progressive events, followed by a clear activity pattern. And again, as you go through the experience with these immune modulating drugs, you see the diversity of different kinds of responses that can be seen.

What I'm going to describe to you here are two approaches that we've taken. First, in the next few slides, I'll talk a little bit about a clinical approach to
[04:01:00] characterization of patterns we see. In the second half of the talk, we'll talk about how we've taken a statistical approach to try to further describe, with greater granularity, the kinds of drivers that might be behind these non-classical response patterns.

Here's one pattern approach that we took, which is something that had been mentioned earlier this meeting. That is, what would happen if we actually re-baselined a patient following resist-based progression while they were on cancer
[04:01:30] immunotherapy? This comes from a atezolizumab trial in non-small-cell lung cancer. These are patients on the atezolizumab arm alone, and when they experience resist-based progression, we re-baseline those patients, and then continued to follow, per protocol.

You'll note here that when you do that, approximately 50% of the patients that had
[04:02:00] already shown resist-based progression will go on to show at least some level of radiographic decrease in their measured tumors, further suggesting that there is a great deal of complexity and a great deal of potential approaches we can take as we try to describe these non-classical responses.

As we look at clinical patterns of non-classical activity, we've tried to break them up
[04:02:30] into some initial descriptors. For example, the classic spike pattern is that that was well-described from the original ipilimumab experience, and detailed in the original IRC paper. But we clearly recognize that there are many more patterns than that. We've seen patients that have sort of an oscillating, waxing and waning pattern of anti-cancer activity, but we've also just started to describe, perhaps, an even bigger class, or a larger class of patterns that we've seen from the atezolizumab
[04:03:00] experience. For example, you can have any sort of a pattern with your sum of longest diameters, but of course, resist would note that appearance of a new lesion would immediately define progression, and so that, itself, would be an individual pattern that we can see with cancer immunotherapy.

Similar to that would be the non-target lesion progression, despite an SLD stability,
[04:03:30] or decrease. What does this look like in actual data? This is data now from our patients treated with atezolizumab in a lung cancer study that would be POPLAR on the left. IMvigor 2-10 is our metastatic bladder cancer study, and you'll note that in this group of patients, these are patients whose first progression is in target lesions, and that progression later reverts, meaning that whatever events led to progression, it has now, with a subsequent scan, no longer qualifies for resist
[04:04:00] progression. You'll see the pattern there is quite diverse. You have the classical pattern of patients who have radiographic progression, followed by dramatic tumor decrease, but many more of these patterns are different. Some of them just represent further stabilization of their disease.

Now, to the second group of patients, note that there are a larger number of spaghetti plots on this slide, and that is because this slide denotes patients whose

[04:04:30] progression was only due to the appearance of a new lesion or a non-target lesion progression. Note, again, the many patterns and perhaps the important thing to note here is just how long these spaghetti plots go out on the X axis. And that, of course, is because, despite the fact that these patients had the appearance of a new lesion or the presence of a non-target lesion progression, they went on to
[04:05:00] derive significant clinical benefit. In many of these cases, patients were followed at the time of this data cut for up to and beyond a year, continuing just on atezolizumab single-agent therapy.

When we start to look at these different clinical patterns, as nicely called out by Dr. Rebus, there is a preponderance of the second pattern. That is, progressions that
[04:05:30] exist just by the appearance of new lesions, or non-target lesion progression, in both our lung cancer and bladder cancer experience that makes up about 25% of patients treated with atezolizumab, versus about 5% of patients treated who had a resist-based progression due to target lesion increase, followed by some further decrease in that target lesion progression.

Now, much has been talked about in terms of different ways to try to measure
[04:06:00] these non-classical patterns of progression. Here, what we've described is what we've referred to as immune-modified resist. This is what we had pre-specified in all of our atezolizumab randomized trials, and even non-randomized trials, and essentially, what we did was pre-specify this, using a resist-based measure but based upon the original publication of IRRC, so you can take a look at this at a later
[04:06:30] point in time. These slides are available, but essentially, it is our own translation of IRRC onto a resist type of approach.

What's important to the remainder of this slide deck is that we went on to try to describe immune-modified progression-free survival, using that measure of immune-modified resist. This was done in a pre-specified fashion. This graphic
[04:07:00] shows the general approach that we took to the definition of immune-modified PFS's, essentially if you have a progression and you try to confirm that progression, if it's confirmed, PFS is dated back to the original resist-based progression date. However, if you had a resist-based progression followed by a non-resist-based assessment, you would no longer be classified as a progressor, per IM PFS. Of course, if you have a progression followed by no further assessments, that would be dated as a progression event, just like in resist.

[04:07:30] When we do that and apply it to a randomized trial, such as the POPLAR study, we can look at what the medium PFS for the atezolizumab arm is, per resist 1.1, and that's noted in the central column, or PFS as noted by immune-modified resist. Of course, by definition, immune-modified resist has to be longer than or equal to
[04:08:00] resist PFS. Of course, immune-modified resist PFS would only be longer in the setting of reversion, that is a resist-based progression event on a subsequent scan no longer qualifying for resist progression.

When we take a look at this in the POPLAR study, again, this is a lung cancer study, you can start to see that there are three major populations that start to form. The

[04:08:30] overall median survival on this randomized phase II study for the atezolizumab arm was 12.6 months, and you can see that if you are a patient that had a resist-based PFS event, and your immune-modified PFS event was at least within one scan cycle, so less than two months' difference, that your median overall survival was 7.3 months, and when you had an immune-modified resist PFS event, for which the
[04:09:00] two event dates were greater than two months, now that group of patients has a 12.4 month median overall survival. Finally, for patients that don't experience any resist-based PFS event, their median overall survival is not reached because, by definition, per any form of resist PFS, these patients have not died.

This is what the Kaplan-Meier curve looks like. Note that the far end of the curve is
[04:09:30] not mature, but again, you can see now in graphical form, what that difference looks like. If you have resist PFS essentially equal to your immune-modified resist PFS, or if the two are different. That difference is driven by these non-classical patterns of response. We can look at this in bladder cancer, as well, medial overall survival of 7.9 months in a second- or third-line plus metastatic bladder cancer setting. If your two events are close together, you're looking at a median overall
[04:10:00] survival of six months; if you look

Section 25 of 45 [04:00:00 - 04:10:04]

Section 26 of 45 [04:10:00 - 04:20:04] *(NOTE: speaker names may be different in each section)*

Daniel Chen: ... you're looking at a median overall survival of 6 months. If you look at those events being different, that you have a median overall survival of 13.3 months and again, another 37 patients who had neither type of progression event. Graphically, this is what the bladder cancer data looks like. Again, here you see not only the 6 and 13 month median difference but certainly a very clear separation across these two curves.

[04:10:30] To summarize the overall presentation here, anti-cancer immunity is a very complicated biologic event that can be influenced by tumor factors, host factors and environmental factors. All of these things conspire to create a very complicated set of potential outcomes for patients. If we were to utilize immune modified recist, a system that has been developed to try to better capture those non-
[04:11:00] classical response patterns, you can then take the next step and not only use that system to evaluate response but actually evaluate PFS type events.

Of course, this builds upon the framework that was created with the original irRC proposed system. Results to date, as we look at lung cancer and bladder cancer data with Atezolizumab monotherapy, you can note that the median progression-
[04:11:30] free survival is increased per immune modified recist relative to recist 1.1, that these changes in progression-free survival event time is associated with a different survival outcome. Certainly it's suggestive that these non-classical response patterns are very much, in fact, associated with ultimate survival. Progression events associated with new lesions or non-target lesions are actually a more
[04:12:00] common pattern in the experience with Atezolizumab as compared to the classic spike pattern where a tumor grows and then suddenly shrinks.

Where do we go from here? The analysis of survival and IM Recist, according to lesion patterns, still give us now an opportunity to take the data from the actual clinical trial experience and start to think about how it should be further modified
[04:12:30] to best reflect the survival benefits that we're seeing from these therapies. The importance of non-target lesions and new lesions to clinical outcome yet isn't fully described. That's something we want to take a deeper dive into. One particularly interesting approach to looking at new lesions are new lesions that are specific to lymph node progressions as new lesions. We recognize in immunity that the lymph node represents a special site in terms of an immune response. This particular
[04:13:00] pattern of progression may ultimately result in a different potential outcome. That requires further analysis.

Alternate IM recist PFS definitions are being looked at. In this case we've defined that IM progression event as being one followed by a direct radiographic assessment that shows reversion. As we all know, and as I've already shown you, there are many patients out there who revert but take more than a single cycle
[04:13:30] scan to do that. One of the things that's being looked at for modification is to allow for multiple scans prior to having that reversion of your PFS event.

Finally, one of the things that's probably been the most enjoyable about our efforts at looking at endpoints has been ... This has created a nidus for industry partners; academic partners; and, of course, our partners in the health authorities, to really
[04:14:00] collaborate on trying to together push this field forward. We currently recognize our ongoing efforts where we're active participants. Not only as part of the EORTC I Recist Working Group, of which the FDA is a part, but also within our own efforts for the Cancer Immunotherapy Consortium which is a group of industry working in the cancer immunotherapy space, to ourselves look at the collective data to help with the global effort to try to deliver better endpoints for cancer immunotherapy.
[04:14:30] Thank you very much.

Marc Theoret: Thank you, Dan. I'd like to welcome next Dr. Keaven Anderson from Merck who is going to speak on immune-related response and survival.

[04:15:00]

Keaven Anderson: Okay, we have the right slide so that's a good start. I'd like to thank the organizers for the invitation and all the speakers for the great talks today and yesterday. I'm going to talk about two parts. One is Merck's rationale for I-R recist, and it's really
[04:15:30] based on irRC that we had in some of our data. It's based some on past and recent data. What I have in that regard was presented by, I think, Eric at ASCO working with Andrea Perrone and Robin Mogg. Robin is from our statistical group.

Then I'll talk a bit about survival modeling and have a few comments on that with
[04:16:00] regards to health economics and what we heard this morning on design and what will hear later this afternoon, as well as a possible endpoint approach which may be the topic of the next talk. I think it has various interesting uses.

In any case, as we increasingly see immunotherapies, OS may be increasingly

[04:16:30] confounded as a primary endpoint due to patients crossing over. We've certainly seen this in at least one of our trials at Merck, if not more. The claim here is that PFS will be increasingly important to assess the clinical benefit of immunotherapies. Recent analyses indicate that they may improve OS with minimal or no improvement in PFS. I mean, the presentations you've seen on the correlation
[04:17:00] between PFS and OS are just amazing and are consistent. I-R recist like determination does seem to correlate better and certainly we need a uniform definition of I-R recist and a few have been proposed here.

I will say that the best approach to this is what Dr. Mandrekar presented earlier. You start with a milestone where you evaluate your response and divide patients
[04:17:30] into groups. In these cases I've got a melanoma and non-small cell lung cancer patients that we classified early according to whether or not they were PD per recist, which is the green curve. They're still alive to be evaluated but they are PD. You can see that that green curve is not a good place to be. Those patients die pretty quickly and have almost uniformly bad outcome.

[04:18:00] Then there's the black curve on the top which looks like a great place to be, especially for melanoma in that they haven't progressed and they survive very well. Not quite so well for non-small lung cancer. Then there's an in-between curve that is a progressive disease per recist but non-PD per irRC. You can see already that
[04:18:30] that's a big enough curve to be an interesting group to look at. You can see that it's distinct from both the green curve and the black curve but it's clearly much better than the green curve. That intermediate response level or the PD per recist but not per irRC seems to be a good place to be. That's important.

[04:19:00] This is just a little bit further explanation of what's going on. This is that same melanoma cohort from Protocol 1. You can see on the left the difference between recist and irRC and if you look down at the table down below you've got a median PFS of 5.3 months by IRR recist and almost 4 months longer based on irRC. That
[04:19:30] [inaudible 04:19:27] is a clinically interesting difference for people. You can see on the right hand part of the figure that a lot of these patients were patients who probably were PD on their first evaluation following up but when you look at IR recist on the y-axis you can see that they go way out and do quite well. Down
[04:20:00] below you see that 22% have longer PFS using irRC then recist.

Section 26 of 45 [04:10:00 - 04:20:04]

Section 27 of 45 [04:20:00 - 04:30:04] *(NOTE: speaker names may be different in each section)*

Keaven Anderson: ... IRRC then resist. So 22% is a figure I think that relates to some discussion this morning. This is the lung cohort. Not quite as big a difference between the two curves. Down in the very lower-right hand corner, you still see there's about a 20%
[04:20:30] group of patients who have a longer PFS using IRRC. The medians here, still over a month difference. Not getting to Dr. Bogaert's criteria for "I want to see my kids graduate from college." But nonetheless, perhaps interesting.

[04:21:00] Interestingly, when you look at the biomarkers selected cohort where we saw much higher response rates, there's still just a small difference. Still, almost a 20% set of patients who have a longer PFS using IRRC. Then we have protocol 6. Now here we have the IRO which is the centrally reviewed endpoints in blue. In green we have the investigator-based events which are based on IRRC. That's how they manage patients.

[04:21:30] The nice thing about this trial is the patients were really managed according to IRRC. Presumably the uniform collection of this follow up data was good in this trial in both arms. Ipilimumab vs Pembrolizumab. I knew that product name. You can [04:22:00] see the difference for Pembro or Keytruda is about a 5 month improvement based on IRRC as opposed to about a 2 month improvement for Resist. It really is a little bit of an artifact that you have this big drop early. There is certainly separation [04:22:30] between the two blue curves and you might say it's similar as the difference between the two green curves, but it's certainly not reflected in the difference in medians which is quite different.

That gets into the non-proportional hazards discussion that came up earlier. I think that's the end of Part 1. Conventional imaging criteria seems to underestimate the [04:23:00] PFS for immunotherapy and certainly data gathering across the industry could support the use of IR resist rather than resist to evaluate PFS.

A little bit on long term survival modeling. This is something that I had asked Jun Shui Ma [inaudible 04:23:23] to take a look at, and he's done a really great job with [04:23:30] it. He's done some analyses that were just presented at ESMO this last weekend. It's based on Ipilimumab-naïve advanced melanoma cohort, keynote 06. The recent data cutoff. It established a cure, long term survival model used to approximate the kappelmeyer estimate.

This approach is a little different from the other ones. I'll talk about it a little bit [04:24:00] more as I get into it, but there's a nice paper with implementation and also references to previous literature in the paper noted here. We're starting with two Pembrolizumab arms, 556 patients. One Ipilimumab arm, 278. This is the latest [04:24:30] data. You can see that we haven't reached what we predict to be the plateau here 49% for Pembrolizumab and 35% for Ipilimumab. We're getting closer. Certainly this has implications for how you will design trials subsequently. If you're expecting end points to accumulate more slowly after a couple years.

[04:25:00] Also note that this 35% is different than the 22% that's been published. This may be reflecting additional therapies that are available in this time period that weren't with the original Ipi trial. I think about 40% of these patients went on to other [04:25:30] relatively new therapies. This is kind of an interesting thing that you can see. We did analyses 9 months apart. You can see our predicted survival rate, the cure rate went form 55% to 50% in the Pembrolizumab group over the course of 9 months.

You say "Well that's maybe not stabilized here." But you can also see that there's a wider confidence interval early on with this model, and it's starting to get narrowed

[04:26:00] down here. Similar for Ipilimumab. Perhaps more interesting is that the difference between the two predicted cure rates is almost identical, and that may be more relevant for the ultimate modeling for health economics, is this difference. It's quite stable. The confidence intervals certainly are useful to evaluate this. The model
[04:26:30] appears to be useful to project possible future survival rates. I think the models with splines this morning are very flexible, and I think what you have to evaluate is the complexity of those versus the ultimate simplicity of this.

This model is a two parameter model. This is just a hypothetical. It can have these
[04:27:00] crossing hazards. There's two parameters. One relates to the cure rate. One relates to the force of progression among those who don't cure. For biological modelers, there's a really cool thing - and I won't take the time to go into - concerning the distribution and number of hits and it's implication on outcomes.

[04:27:30] If you're only modeling the cure rates here, and the others don't, which was adequate for this model. This is actually a proportional hazards model. It includes a cure rate and proportional hazards. It gets at what we're used to seeing, but also gets us to the ultimate question of, "Am I going to see my daughters graduate from college?"

[04:28:00] It's a nice model. Related to this, we haven't gotten to this yet. You can model either of these two parameters based on co-variants. This is something I did 25 years ago for the Weibull models that Dr. Latimer referred to in predicting cardiovascular events from the Framingham Heart Study over 30 years. Those
[04:28:30] models here are inadequate. If you try to fit those, they just simply don't fit in the context of a cure. That's why it's important to get to the cure mixture model. I think this Poisson mixture model is what we're referring to here, seems to be particularly useful. There is available software in the reference for estimating these things.

Basically the usefulness, it's stabilizes point estimates for smooth Kaplan-Meier
[04:29:00] curves. You might get a better estimate with the median or the long term survival rate. Confidence intervals are very useful to estimate how accurate, especially the difference appears to be and that certainly deserves further examination. We have been using this now for designs of four trials going forward. One for the follow-on trial for protocol 6 with the idea that events are going to accumulate very slowly at
[04:29:30] some point because of the lowering event rates. But also in trials for adjuvant and neoadjuvant where you expect a substantial number of patients not to accrue. They appear to be very useful for those to more closely model what physicians expect and very simply model it.

I think that's all I've got and we can move on to the next talk. Thank you.

Section 27 of 45 [04:20:00 - 04:30:04]

Section 28 of 45 [04:30:00 - 04:40:04] *(NOTE: speaker names may be different in each section)*

Marc Theoret: Thank you Keaven. We're going to move on to our next talk. I'd like to welcome Dr. Chen from Bristol-Myers Squibb to discuss alternative survival endpoints.

[04:30:30]
Tai-Tsang Chen: Tai-Tsang Chen from BMS. Thank you for your kind invitation. The topic of the talk is alternative survival endpoint. Here's my disclosure. I borrowed this slide from Dr. Urba's presentation in 2013 at ASCO. At the time, only one immunotherapy, well immune checkpoint inhibitor, was approved which was ipilimumab. At the time, we
[04:31:00] show that ipilimumab has a long-term survival effect. The prediction at the time was PD-1 blockade would induce even better survival, long-term survival. If you have a combination of anti-PD-1 and ipilimumab agent or anti-CTLA-4 agent, you're going to lift the curve even further. Eventually if you have a combination or certain sequencing, you're going to lift this to a point that, as many presenters have
[04:31:30] indicated, it's going to make survival trials extremely difficult. This shows a promising future for cancer research, but also it shows a lot of challenges for us.

The focus of this presentation is- Many, many previous presenters have already focused on the earlier endpoints that could be used to likely to predict clinical
[04:32:00] benefit, essentially we're talking about over a survival year. The focus of my presentation is going to be on survival again, but is not on only survival but I would like to focus on two alternative survival related endpoints. I understand that the ultimate goal is to find [inaudible 04:32:14] point, but before we get there I would like to maybe come back a little bit and look at over- and survival again.

These two survival endpoints are milestone survival and restricted means survival, and obviously these two endpoints were mentioned a couple times in previous
[04:32:30] talks already. So what are those? Milestone survival is defined as the Kaplan-Meier survival probability at a specific time point. We define this a priori in a study design, or you can do this post hoc in the study report such as two year survival rate, three year survival rate. It doesn't have to be survival, it could be PFS or any time-to-event endpoint, as long as you can come up with your Kaplan-Meier curves, you can come up with a milestone survival, or milestone PFS, for example.

[04:33:00] In terms of the relative treatment effect, this can be measured by the difference of two milestone survival, or the ratio of two milestone survivals. Regarding restricted means survival time, essentially this is the area under the Kaplan-Meier curve. It has a very nice interpretation, it is the life expectancy for restricted or truncated duration. We have the long tail. In order to calculate the area under the curve you need to have a closed curve, otherwise you cannot calculate the area under the
[04:33:30] curve. Therefore you need to cut it somewhere, and that truncated duration would be for example, a patient's life expectancy for the next 24 months is going to be 19 months. It has a very nice interpretation to that. Also it can be measured by the difference as well as ratio.

Many of you have seen this already. To set up my next slide, I'm going to talk about
[04:34:00] this just very briefly. Also presenters also have already presented this. Panel A on the top left corner, proportional hazards model, these two curves are generated using exponential distribution. As we know the exponential distribution shows you constant hazard, when you have two exponential distributions to represent two curves, then you have a proportional hazards. Panel B shows you long-term

survival. In this case, very similar to Dr. [Kenneth 04:34:20] Anderson's presented. This is also a proportional hazards model, except that I incorporated the cure and the cure is a little proportional to each other. Mathematically, when they're
[04:34:30] proportional to each other in this sense, it's actually on a locked scale.

Panel C shows you delayed clinical effect. The way that I generate this curve is I assume a piecewise exponential distribution during the delay obviously hazard ratios go to one. Afterwards, from the time the separation, you have a proportional hazards model. Panel D is the most complicated one. It incorporates [inaudible 04:34:54] effect, the long-term survival effect, and indeed this is the phenomenon
[04:35:00] we see in most of the immunotherapy trials. Let's look at how good they are compared to each other. This is just one of many scenarios based on simulation. I need to have a benchmark, so I designed a study so that the traditional overall survival endpoint, the power is going to be maintained at 90%. And here are the design assumptions, I want to have minimum follow up duration of 24 months, meaning that from the time the last patient came on to study until the time I
[04:35:30] analyze the data, I want to have at least 24 months of follow up.

The power based on the [inaudbile 04:35:33] test, the overall survival endpoint, is going to be 90%. This end point. False positive rate alpha is 5%, two-sided, and control-arm I'm assuming exponential distribution with a median of 12 months, just for simplicity. As I mentioned, because I'm using overall survival as the primary endpoint to serve as a bench mark, I'm assuming for the 4 models- By the way
[04:36:00] these 4 models represent to the curve that I just showed previously, A, B, C, and D. PHM stands for proportional hazards model. PHCRM stands for proportional hazard cure rate model. NPHM represents non-proportional hazards model, which means that when you have a delayed treatment effect in there. The last one is the most complicated one that includes both cure rate [inaudible 04:36:23] but also cure, so that's non-proportional hazards cure rate model.

[04:36:30] This set of hazard ratios will give you power of 90% for the overall survival. When there is a cure or a long-term survival effect I'm assuming .1% in the control arm, and when there's a delay I'm assuming 8 months delay. Remind you that the median in the control is 12 months. Let's take a look at this. You will notice that when you have a proportional hazards model, that is the first one and the second one, first column, second column. [inaudible 04:36:58] test is the most powerful.
[04:37:00] It's 90%, in this case, milestone survival not so much. Why? Because it doesn't take into account the totality of the data. You look at just one time point. By the way, I'm looking at 2 year milestone, and also two year restricted mean survival. The reason for that is because I restrict my minimum follow up of 24 months, I want to make sure that the estimate is robust. I don't want to go beyond that because the curve will be fluctuating, because I have a lot of patients who will be censored due to end of the study, for example.

[04:37:30] However, when you get to non-proportional hazards model and non-proportional cure rate model, you'll notice that the milestone survival, if you design the study based on the 2 year milestone survival, then you start gaining power. Obviously this

is one of the many scenarios that you may encounter. I'm not saying that this set of numbers is going to apply, can be applied to all the scenarios and all situations. So here's the question then, one thing that you notice is that overall survival is pretty
[04:38:00] good. I personally believe that [inaudible 04:38:02] test is a very, very good test. Overall survival as we said is a very. very good endpoint. So why do we still want to consider other endpoints? In this case, why do we consider other survival endpoints?

We know, as I mentioned, non-proportionality in hazard ratio among [inaudible 04:38:18] agents is a fact, we have seen this multiple times. As I mentioned, the scenario that I mentioned is just one of the many, many scenarios that could
[04:38:30] potentially happen. By considering alternative survival endpoints you may gain power for the study. Secondly, the need for describing potentially time-dependent treatment effects. Traditionally we design a study assuming exponential distribution. You may or may not be aware that the reciprocal of the hazard ratio is essentially the ratio for medians.

[04:39:00] You can present hazard ratio, but there's a need to present the absolute treatment effect by r, you don't have that number. So what do people do? They report medians. Why? Because this is how the study was designed. If everything goes as planned, of course the ratio of the medians is going to be the reciprocal of the hazard ratio, so presenting median survival is very meaningful. We have all been
[04:39:30] taught mathematically that median is good because it is immune to outliers, but the outliers we're talking about here are long-term survivors. So now the advantage becomes the disadvantage. So what we need to do is can we come up with some other measures that can capture that number?

Finally, the unpredictability of the analysis time based on overall survival. Even if you take that into account, you build in the survival time, you have the cure rate,
[04:40:00] you have the delay treatment effect, there's no guarantee that you're going to get the-

Section 28 of 45 [04:30:00 - 04:40:04]

Section 29 of 45 [04:40:00 - 04:50:04] *(NOTE: speaker names may be different in each section)*

Tai-Tsang Chen: There is no guarantee that you are going to get to the end point or get the number of events that you need and the time, and that was a very real example that when I was the lead statistician for Eribulin the front line medicine went on trial. Patients were treated with Eribulin plus minus dacarbazine, the trial was designed for three years and it took five years for us to conduct a trial, great news for patients. As a sponsor or cancer researcher it was extremely difficult because we do not know
[04:40:30] whether or not the drug is working but we understand that the hazard is dropping. We were experiencing one event per month, pretty much, toward the last two years.

So here's an example of the impact on the specification of long term survival effect on the trial duration. The study was designed ... As you can see there's a set of assumption at the bottom here. I designed this study with a median control again is

12 months, and I'm assuming this time six months delay. And post delay I'm
[04:41:00] assuming hazard ratio of point five. When I design a trial I design for 12 months of cure rate. I want to follow a patient for 24 months and there is no cure rate. Okay, if everything goes as planned I'm going to finish my study in three years.

Let's say that you in the control arm you now have 50% cure rate. And I'm using the term of cure loosely, you know, long term survival. Now what happened is that it'll take you 12 months to enroll patients, but your follow up duration is going to be 51
[04:41:30] months before you hit that number of 251 patients. And it turns out that the study now based on studies that I'm going to have 251 events out of 345 patients that I've claimed to enrolled. But it so happens that you actually have a long term survival rate of 30% in control, you never finished your study. Why? Because you don't have enough patients at risk for it to reach 251 events.

Now I want to shift the gear to reporting in terms of how to report a timed event
[04:42:00] endpoint. Again I'm illustrating this based on a simulation. Conventionally we report clinical trial results using the following summary statistics. This is based on simulated data by the way. Hazard ratio 95% confidence interval. In this case it's .88, confidence interval ranging from .76 to 1.03. And then report on the median survival as well as the confidence interval. In this case experimental treatment is
[04:42:30] 9.4 months, control is 11.4. And the log rank is .1. So these are numbers that you see all the time in the literatures and how we report the clinical trial outcome. Here are the issues. Hazard ratio indicates that experimental arm has a better performance because hazard ratio is .88 regardless whether that is positive or not. The log rank of .1. It also shows that a consistent improvement in the treatment effect relative to the control, because it's a constant hazard ratio, so it's .88. By
[04:43:00] your meter survival is better in the control arm.

So how do you put this in your head? It's difficult to interpret it. So automatically you know that the curves cross somewhere. But where? That is very important information that is not being show here. In your head this is what the hazard ratio means. Constant hazard across time, no change. If you plot this differently, then
[04:43:30] you can see this and this is going to offer you a very different information. I'm plotting this ... You'll see that there is the black curve. This is the hazard ratio over time and the shaded area is the 95% pointwise confidence interval. Here I also added the dotted lines that are representing 95% simultaneous confidence interval bend, and when you are thinking about the confidence interval bend you can interpret it thus. For example, if you are testing the same input multiple times usually you need to take a penalty right? So it is very similar to that concept that if
[04:44:00] you look at this confidence interval on the same input multiple times you need to widen its confidence to ensure that the entire curve actually has 95% coverage.

So this is what it is. What it tells you is the drug starts to get better at around eight months and that you will observe a constant hazard ratio after 13 months. But when you see that when the hazard ratio curved, crosses one, that doesn't mean
[04:44:30] the survival curves cross at one. That means that at that time the curve in the experimental arm starts catching up with the control arm. So what you should have

in your head is that you know the curves probably cross somewhere between ... Definitely sometime after eight months.

This is another way of presenting this information. But instead I'm showing you the difference in monitoring survival over time. And again you have the shaded area of
[04:45:00] 95% confidence interval, and the dotted line is the confidence interval bend. And the black line means that it's a pointwise, over the estimates. In this case just so happens that 24 months you see that confidence interval actually excludes zero. So this is a different way of looking at the data. And in this case it shows you very clearly the curves cross at 13 months approximately.

Finally the restricted mean survival over time. For normal people you will see that
[04:45:30] the curve, the life expectancy for example, should be close to the dotted line. Okay. The life expectancy for people who are sitting here in the next 24 months will be approximately 24 months. Therefor you see a curve. It tells you that the blue curve is the control arm, at the beginning, probably at 12 months, is slightly better than experimental arm. But when it gets to 24 months, they are pretty much the same. This is a different way of presenting the restricted mean survival. Instead I'm showing you the difference. So you also see that the difference actually starts to
[04:46:00] shift after 13 months and that's when the curves cross.

This is the underlying Kaplan-Meier curve. This curve also shows you one thing that nothing that I just presented showed you. That is the total follow up. The reason why I calculate everything based on 24 months as I said, I want to have a robust estimate. That's why I stop at 24 months. But without this curve, you don't know
[04:46:30] how long the patients have been followed for this study. All patients who are censored at the end, those are the patients who are censored due to the end of the study. So those are the administrative censors.

So in summary, although survival remains a very good endpoint, under certain assumptions, alternative endpoints such as milestone and restricted mean survival are very helpful because they have predictable length of study duration. So they are time driven. There are also survival endpoints. And they allow both the relative
[04:47:00] and absolute measures of true mean effect over time as well if you want to do it that way. They are more powerful under certain patterns of non-proportionality and hazard ratios. So in my opinion it is important to report milestone survival, restricted mean survival, and change in hazard ratio when you have a clinical trial that you need to report. In my opinion you need to have a set of summary statistics that can help you redraw the Kaplan-Meier curves in your head. If you cannot do
[04:47:30] that you may have a wrong set of summary statistics.

Finally there's a list of references. Thank you.

Marc Theoret: Well thank you Tai, I'd like to welcome all of the presenters up to the stage here
[04:48:00] and I would also like to welcome Dr. Kun He from FDA who will be participating as a panelist as well.

[04:48:30] So I'd like to thank all the presenters for excellent presentations. And what we heard was two different kind of approaches. One, multiple different presentations using tumor response based endpoints and really looking at alternative tumor response based endpoints, to potentially capture unique considerations for
[04:49:00] immuno-oncology products. And most of the date presented have been from really one class of immuno-oncology products, checkpoint inhibitors. Then we also heard about other types of ways to measure clinical benefit endpoints such as overall survival.

I did want to just start the panel discussion with one of the questions from the folks participating on the webcast. And the question was ... This will be for Dr. Mushti.
[04:49:30] For the analysis presented, was the PFS OS analysis confounded by either crossover after progression, or post-treatment use of immuno-oncology for patients progressing on the control arms?

Srisha Mushti: To answer that there was only one trial which allowed for the crossover. But the data that we considered was based on the interim analysis which was before the
[04:50:00] crossover took place. So the concluder, no crossovers that we were including in our analysis.

Speaker 2: Okay, thank you. Going-

Section 29 of 45 [04:40:00 - 04:50:04]
Section 30 of 45 [04:50:00 - 05:00:04] *(NOTE: speaker names may be different in each section)*

Srisha Mushti: Including in our analysis.

Speaker 2: Okay. Thank you. [crosstalk 04:50:04] Go ahead.

Male: Because when I was looking at your presentation which was very interesting the correlation seems to be better whenever we have a higher PFS and a higher OS. If you have lets say a response rate of 5% it's hard to impact survival. If you have a response rate of 75% it's doable, it's very likely they will achieve survival and the
[04:50:30] lack of correlation in your curves seem to be in the lower end. If you discard the ones who respond less than 15%, you may have a better correlation between PFS and overall survival.

Female: I think 15% is also explored by Dr. [inaudible 04:50:48] and according to her conclusions, she mentioned that even the associations didn't seem to be pretty strong with the 15% control, a 15% cutoff.

Male: Thanks.

[04:51:00]
Marc Theoret: So why don't we go to microphone three, we'll start with some questions from the audience.

Speaker 4: Thanks. Those were great presentations. Dr. Mushti, for the analysis of PFS, based

on non-target or new lesions, how are those two categories handled? Or was it just on target lesion worsening?

[04:51:30]
Srisha Mushti: The progressions are definitely based on all three. The ones that the modifications that we have included would majorly impact only the target tumors.

Speaker 4: So you still considered non-target progression-

Srisha Mushti: Yes.

Speaker 4: Okay. [crosstalk 04:51:39]

Srisha Mushti: And new lesions too.

Speaker 4: And then for the response metric, Dr. Gao, the analysis, did you look at all at just the AUC of the negative inflection at the beginning? Or did you just look at the depth of it?

[04:52:00]
Xin Gao: We looked at the depth of that, so there are ones set of the patient that has no reduction at all. But we used the smallest percentage to capture, but we use positive value of the AUC compared with the negative value of the AUC the reduction of the depth is reflected as the positive value in the AUC.

Speaker 4: Did you consider at all looking just at the negative AUC?

[04:52:30]
Xin Gao: I think we considered both the positive and negative values.

Speaker 4: There may be some [crosstalk 04:52:38]

Xin Gao: Do you want to ask?

Speaker 4: Yeah.

Speaker 6: What you are asking is if we could define only those who had any reduction as response. We didn't want to look at that and also you see if you look at that there
[04:53:00] could be some patients, even then we would have to cut off at one mirror or something like that to look at the AUC, so we just picked mirror and whatever was the area, so we could use all the data that we had.

Speaker 4: Okay, thanks.

Marc Theoret: I think before we get to the next audience question, one of the things that is very much apparent from many of the presentations, looking at the different IO products at least for the checkpoint inhibitors, is this phenomenon where some

[04:53:30] patients will have this resist based progression for example, and go on to have either non-PD based upon target lesion reductions, or in a small number of cases as Tony pointed out, actual reduction below what would be considered a subsequent, resist response.

[04:54:00] I guess one of the major considerations I think we all have to think about, I'd like our panelists to think about, and respond to in a bit, are how much are these evaluations for alternative efficacy endpoints looking at these responder analysis, how much are they limited based on the lack of control arm, in those settings? Because even with the IO products there's certain safety factors that are basically in place so that we're not exposing patients to the risks of potentially ineffective
[04:54:30] therapies, to a great degree. Such as patients shouldn't have rapid progression for example, or they shouldn't have decline in performance status to continue on treatment past progression as well as other potential criteria that many sponsors have put in place. I just want to put that out there and I think we'll have some discussion of that because I think it is an important point about moving forward & think about the limitations of the information that we have, and we may have the
[04:55:00] ability to look at that from what we have or collect information prospectively as we design trials to address their potential limitations.

Why don't we go to microphone two here first.

Speaker 8: [inaudible 04:56:12] independent. We should always think outside of the box and try to learn from other areas and one thing that occurs to me is that in drug
[04:55:30] development in clinical trials in areas such as cardiology, or diabetes, survival may not be necessarily the only endpoint. And very often we see composite endpoint where mortality certainly is one of them. But then you have other events. Stroke, MI, schema, you name it. Now our PFS analysis in oncology, they do contain mortality obviously but they also contain progression as an event. Is there anything
[04:56:00] that we might have learned or that we should learn from other therapeutic areas as to apply to our analysis in the search of an intermediate approach to long-term survival?

Male: I'll take a shot at that. I don't know what other therapeutic areas you may be
[04:56:30] referring to. I'm certainly not an expert in non-oncology development. Even within oncology we know we have other measures that we haven't taken into account yet here. An example of that of course would be all of the work around biomarkers.

Biomarkers can be difficult if they're tumor based. Most of the cancer immunotherapy biomarkers to date have been tumor based, but if we could develop blood-based biomarkers that would be a way to both start to tease out
[04:57:00] whether patients really have an ongoing biologic response to cancer immunotherapy as well as potentially be a read-out for benefit. And while the immune based markers have not shown specificity for benefit to date, there has been an inkling of a signal from cancer, blood-based, biomarkers. One possibility as we think about how to dissociate patients receiving benefits from cancer immunotherapy despite all of these non-classical patterns, will be to look at things

[04:57:30] like circulating biomarkers for individual tumors, or even tumor associated DNA. Those would be alternative methods from within our own field to try to dissociate which patients are going to benefit, and as you asked, who should be getting continued therapy.

Male: [04:58:00] Even we don't detect the composite endpoint so far, but if you see our approval process in a sense, we look overall performance or look a variety of secondary endpoints subgroup, so we always set a totality of the trial results. In that sense, we conceive everything.

Male: [04:58:30] [04:59:00] A couple comments. I have found it useful in the past to look at different components. Are progressions due to new lesions, are they due to increase in size or are they simultaneously occurring? Certainly that's useful. Another thing that I found interesting that is really only anecdotal, is if you're looking at lesions, metastases in different organs, they may be behaving quite differently as well. And then finally, I feel like going from 20mm to 30mm may be quite different that going from 100mm to 150mm even though they're the same by resist criteria. They're plenty of directions you can go in to try and understand things better.

Speaker 2: Okay, microphone three.

Male: [04:59:30] [05:00:00] Thank you, thank you all for these great presentations and especially Dr. Gao & Dr. Mushti for sharing the FDA's analysis. I think that's exactly where we need to go, that we need to explore in the data mining way, kind of new intermediate endpoint. Specific question for this intermediate endpoint proposal that you shared, first thing I saw that you considered new lesions within one year as basically ruling out response? How does this fit together with what we always hear, like you can't have, for example new lesions in the beginning. So for example did you try to exclude those new lesions in the first few assessments maybe? And a second question because I understand you need kind of a one year follow up, did you compare the correlation you get with milestone for example survival at one year, and how does that compare to that?

Section 30 of 45 [04:50:00 - 05:00:04]

Section 31 of 45 [05:00:00 - 05:10:04] *(NOTE: speaker names may be different in each section)*

Speaker 1: And how does it compare to that?

Speaker 2: [05:00:30] We have three factors, non-target lesion, target lesion and new lesion. Using the online criteria of a target lesion, they're about 20% of the responders, and then if we include in the criteria of non-target legion that will reduce about two percent. And if we include additional criteria from new lesion that will decrease three percent. Not a huge decrease or huge effect on the criteria yet.

And regarding to the overall survival, the median in the overall survival, the median of overall survival in the meta-analysis is about one year.

Speaker 1: Thanks. And did you also try to compare the correlation you get with just taking, for

example, percent of patients alive at one year after starting treatment? How much
[05:01:00] correlation do you get with OS using just this milestone survival, because I see you have this one year follow-up period.

Speaker 2: No, we didn't look at that in detail, but that would be interesting work in the future.

Speaker 3: I just have a follow-up comment on that incredibly interesting analysis that the FDA has done here. So you essentially have optimized some of the major criteria that
[05:01:30] we see for cancer immunotherapy, but it's optimized specifically for the agents tested.

And so I wonder if, as we think about these criteria for the future, whether as part of your analysis you can start to categorize the different patterns you see, and you've optimized it for the existing agents. But if you took now a new agent, that had a different balance of these factors, is there a way to quickly pivot to adjust the approach for an agent that has a very different profile?

[05:02:00] As an example, what if you had an agent that, through its biology, it literally generated inflammation so you immediately saw new lesions?

Speaker 2: Right now, we consider the time of one year, and definitely we will explore a [inaudible 05:02:17] time point in the future. We would like to shorten the data, like six months. Right now, we're composing this component in point, so we are
[05:02:30] given similar weight so when we make this [inaudible 05:02:32] point, they have to satisfy all the three criteria, non-target lesion, target lesion and new lesion. But there may be a lot of ways to combine this criteria, and then combine these three components and then define the criteria. That will definitely be our work in the future.

[05:03:00]
Speaker 4: I have a question for Doctor [inaudible 05:02:57]. You talk about reset the baseline. Could you talk about it a little bit more, and try to redefine the patient population after you reset the baseline?

Speaker 5: I think the data that you're referring to is something that was presented this year at ASCO from our POPLAR data set, and you're right, that's what this sought to do.
[05:03:30] The POPLAR trial allowed investigators to assess a time of radiographic progression, or RECIST progression, whether patients would have ongoing benefit. If the investigator assessed that the patient had evidence of ongoing benefit, they were allowed to continue on the Atezolizumab treatment, despite that RECIST-based progression. And so the re-baselining effort essentially took the existing lesions, including the lesions responsible for progression, and now reset that as your new
[05:04:00] SLD. And then the waterfall plot you saw was the percent change on subsequent scans for those patients, based upon that reset SLD.

Speaker 4: Okay. The following question is for the [inaudible 05:04:19]. In terms confirmation,

[05:04:30] if the change from baseline is already 100%, do you think it's still necessary to have a confirmation?

Speaker 5: I can start with that?

[05:05:00] The patterns of pseudo-progression or non-classical response are incredibly varied. And yes, we have examples where there is no percentage change that I've seen yet that qualifies as definitive biologic progression. We have patients that have surgical evidence of clear pseudo-progression, because you go in and you resect this massive lesion now and it's all lymphocytes with very, very few viable tumor cells remaining. So I don't think, as a field, we're in a position yet to say there's a particular percentage that defines true biologic progression.

[05:05:30]
Speaker 6: So just what I would follow-up on that, if you're re-baselining, your treatment effects may be very different on the re-baseline, if you're comparing two treatments. Do you have any thoughts on that?

Speaker 5: So of course, we're not actually proposing re-baselining, we only showed that one data slide as a consideration as we think about paths forward. None of the other work I showed actually involved re-baselining. But you're right, there are a number of complexities that come with the re-baselining effort. It's interesting, if you take

[05:06:00] the very same patients on that waterfall, and you look at that waterfall by re-baselining versus just using the original SLD, you'll imagine the magnitudes of changes are different. Because the baseline you're comparing to calculate your percentages are different. And if you then try to take a look at that versus a control arm, you're going to have to correct for those differences.

[05:06:30]
Speaker 6: So it's more of an exploratory and explanatory analysis, as opposed to something that you could use for regulatory purposes, it sounds like.

Speaker 5: It's an interesting way to put it.

Speaker 6: I think it's very interesting, don't get me wrong, I think it's extremely interesting.

Speaker 5: I think you're right, that at the first instance here, what we're doing is trying to describe a biology. And I think the re-baselining effort clearly defines an a biology that is present. That being said, whether we can successfully or whether we even

[05:07:00] should include it, as part of regulatory end points, we just don't know. I think we're still very early here, I think there are large efforts intended as part of these collaborative efforts, to bring together the data sets, and I think that's ultimately what's going to be needed to show whether re-baselining or any other approach is ultimately valuable in a regulatory sense.

Microphone three.

[05:07:30]
Speaker 7: Thank you very much, [inaudible 05:07:28] from John Hopkins, I'm a statistician, and first off thank you so much for the excellent organization of the conference and the excellent talks from all speakers.

So my question is maybe particularly to Doctor [inaudible 05:07:49]. The majority of this session's talk is that we try to basically, my impression is that we try to
[05:08:00] capture the tumor growth dynamics in some sense, such that it can be somewhat more informative in this new check inhibitor regiments. So I just wonder that, when we first develop a RECIST criteria to categorize or to capture the tumor growth and
[05:08:30] kinetics for a cytotoxic agent, is there any lessons or experience that maybe you can share with us that, how you come up with this relatively simple criteria which is being used here. I believe that for cytotoxic agents, the tumor dynamics are so complicated, and it can be described or modeled much more complicated instead of here, so I just wonder that ... Because, as a statistician, as I was just sitting there I
[05:09:00] could think of a lot more complicated way to model the whole trajectory. But just, why this? I believe there must be some practical consideration that's important.

Speaker 5: Thank you.

Speaker 1: The original development of RECIST was essentially trying to convert WHO into a
[05:09:30] uni-dimensional method. I wasn't part of that, by the way. After that we tried to further modify it to remain very simple, so you want a methodology that can be done in every hospital, in a trial. When you collect those databases and you start
[05:10:00] going through them, I think Sumitra expressed that the best. The quality is limited, there are holes in those data, the method of ...

Section 31 of 45 [05:00:00 - 05:10:04]

Section 32 of 45 [05:10:00 - 05:20:04] *(NOTE: speaker names may be different in each section)*

Male 1: Limited their holes in their data. The method of measurement varies. You see all kinds of issues and I'm very interested in the work that was presented by FDA where you take all the data up to one year, which of course is not available at every time for the patient so you cannot use it to instruct the future treatment of the patient but you can certainly do much better in optimizing to our survival. That's
[05:10:30] for sure but we come from a very simple system I would say to try and have a broad ...How you call this?... A threshold to be able to compare trials from various origins and trying to now modify it to become something that is usable as a
[05:11:00] reliable, almost surrogate in a new setting and that's a challenge and I find it very interesting myself. We've also done modeling that gives better results if we can make bigger compound models but the original idea of resist was let's say who was a responder at the moment itself at 4 weeks, at 8 weeks, at 12 weeks, so it's come a long way.

[05:11:30]
Marc Theoret: So I realize I've been very deficient on the right side of the room. I saw a lot of jumping around so please why don't you go ahead.

Dr. Phan: Well, thank you and actually I am on the left side.

Marc Theoret: My right. My right side. [Laughs]

Dr. Phan: I'm Giao Q. Phan, one of the oncology medical officers at Saber so Mark you are allowed to neglect me. It's in the family. I am very interested in the questions ...The
[05:12:00] presentations about Dr. Anderson. Did I get it right? Dr. Anderson? Yeah. At first I apologize that I didn't attend your full presentation. I just didn't get a chance to get away as I wish but that was very interesting about the progression for survival by traditional resist and there immuno-IR response criteria. So you see by immune
[05:12:30] related criteria, you see the progression for survival longer than the traditional which looks good and feels good too but I have 2 questions.

First of all, could you please comment on: In a higher granularity, what makes that difference? My understanding that traditionalist two-way is mostly good
[05:13:00] ...Traditional resist... When you have a new lesion, it's got progression but by IR you are not so I suspect that you will see more of partially responsive, more than the traditional one, so I would like to know a little bit if you have that data in depth about the correlation between the complete response. The difference? Is that a complete response or partial response stable or whether that related to survival is
[05:13:30] number one question? Number two is the resist that looks good, I wanted to know whether the patient feels good, so if you have any data about that or PRO-Patient Reported Outcome with that? Thank you.

Keaven Anderson: Yeah. Thank you. I don't know if we have specifically addressed the correlation of
[05:14:00] PRO with the outcome but we do have PRO data that is very favorable within the studies and that's been published so we were very encouraged by that. It was a less certain hypothesis coming into the trial than you know what we were looking for and PFS and OS. Eric Rubin may want to comment but my understanding would
[05:14:30] tend to be more PR than CR that would be the differences. Now we've heard descriptions of Pr's that may effectively be CR's here but in terms of the biology by many people but I certainly can't comment on that but I would say it's primarily through PR's. Did I address your questions?

Dr. Phan: That's right. Thank you.

Keaven Anderson: Okay.

Marc Theoret: Microphone two.

JSK: Yeah. Askavos JSK. Wonderful presentation and just one step forward as I had
[05:15:00] hoped from the earlier session we would go today but I just want to push this one more step and ask the potentially obvious question. We have now made the link between the biology and the way we want to describe response. It isn't perfect but I think we have taken a step forward to beyond the conventional criteria that were developed for chemotherapy and to describe what this biology is and how we can
[05:15:30] capture it. There is still more work to be done in terms of landing on the right

system but I think what was just said here about practicality of the criteria. It has to be practical for routine clinical use. If it isn't, if it's too esoteric, it's very hard to implement. So with that in mind, what do we think is needed burden of proof for something like this to become acceptable, credible on a regulatory level, not just in
[05:16:00] the hands of a clinician that who does the clinical trial and is looking for an exploratory measure but actually elevate this towards something that is more of regulatory ability? ...

And when I say that, really in the sense of describing biology and capturing clinical benefit more accurately than what we have done before and frankly PFS and
[05:16:30] regular response, historically, has never undergone the same scrutiny that we are applying with what we are doing here so the gold stand in itself isn't that golden so having said that ...You know... What is the opinion of the panel? What do we actually need to show so this is gaining some credibility?

[05:17:00]

Male 1: For me, personally, the thing you need to show most is that if you go beyond what we today call 'resist progression' that the amount of dramas is limited because we know we're going to have situations where treatment is extended for patients but they are true progressors from the start and if you can quantify that and you can
[05:17:30] show that the risk benefit during that period of time is positive then, for me, that would prove that it is a valuable approach for me.

Male 3: I would certainly hope the FDA would comment on this but I'll provide my own perspective on those questions first. When we look at this effort, I certainly hope that what we're not doing is something that's 3 or 4 steps behind what's already
[05:18:00] happening in the clinic. That is we don't need more ways to show that PD1/PDL1 inhibitors work. It's helpful but we already have traditional methods of showing those kinds of benefits and sure there will be opportunities whether it's once you start going against stronger standards of care that you may need to be able to better show this but I think the biggest need for us in the future, again, comes back to the fact that we have 800 combination cancer immunotherapy trials already in
[05:18:30] the clinic today and what we don't want to be is always 3 steps behind where we're redeveloping new tools for something that's already shown clinical benefit.

And so my hope is to address your question that we get to a place where we can show the patterns now that are important, that we can use existing databases to link them to survival and as you mentioned, create something that is practical
[05:19:00] enough and flexible enough that can then be applied to the next generation of therapies that are quickly coming and again, from my own perspective, I think the idea of allowing for something like reversion of a PFS event is the kind of tool that's very flexible ...That you can imagine no matter what you bring to it... None of us are going to argue that a patient that shows progression followed by some level of non-progression for an extended period of time isn't clinical benefit.

[05:19:30]

Male: So I would definitely echo that point about kind of needing, though, to look back to

look forward because it, as Dan had mentioned, this has not necessarily been an issue for the recent generation checkpoint inhibitors so just of all checkpoint inhibitors, the four that are approved, there's been 16 new or expanded indications of which 9 were accelerated approvals ...So of those 9 accelerated approvals, 7 of those 9 were based upon objective response rates in duration of response so
[05:20:00] clearly as a tool that has been used from regulatory...

Section 32 of 45 [05:10:00 - 05:20:04]

Section 33 of 45 [05:20:00 - 05:30:04] *(NOTE: speaker names may be different in each section)*

Male: Clearly is a tool that has been used from regulatory perspective to identify anti-tumor effects that are reasonably unlikely to predict clinical benefit. Some of the trials have gone on to confirm clinical benefit for some of those accelerated approvals. It's not really ... I think we always need to push this forward and not necessarily say what is necessary from a FDA perspective. What is necessary from
[05:20:30] the scientific perspective? What is necessary to demonstrate what we're intending to demonstrate. I think we always need to be very critical.

I thought Dr. Schwartz gave a nice presentation about how much can we rely on just the, at least from our perspective, data sets, case report forms, when we're trying to analyze progression and degrees of progression if that information is not necessarily captured in a way. Do we need to go back to the actual images and do a
[05:21:00] lot of these analysis which I believe some members on the panel has already done for their excellent exploratory analysis for trying to capture some of these unique benefits. One of the points I brought up much earlier and I would like the panel to talk about this a little bit is how much are some of these analysis limited by not having a control arm because those are kinds of design issues and I think this will
[05:21:30] lead into some of the next session that could be addressed if not already addressed in certain trials.

Tony: With a fine of respond to benefit, obviously we need to applied the same way to the experimental arm and the control arm, that's obvious. If you have an effect on
[05:22:00] a series of patients, for example the tumor shrinks, you may conclude, as the FDA does, that that tumor shrinkage is meaningful and then you give accerlerated approval. Not always you need a control arm. If it's a shrinkage that's long lasting, you may not require a control arm because if the control arm get the same that will be what we would all be using.

[Tom Bergert 05:22:20] said the thing that was very powerful at the beginning of
[05:22:30] his talk which he said progression by [inaudible 05:22:29] was not intended to tell physicians what to do. I think that empowers us a lot more to look beyond progression and as [axel 05:22:39] and others had defined to confirm progression in patients where it makes clinical sense to do so, not prolonged futile treatment. Allow something that we know, it's possible that may happen. If we take those
[05:23:00] patients who progressed and responded, we'll look at them retrospectively, I think all of us we would conclude that those patients benefited. It wouldn't need a whole day of discussing about pseudo progression and how we capture it. It would look retrospectively. The problem that we have is that we all want to go very fast and

we want to know by 12 weeks did that patient respond or not, is that curve better than the other.

[05:23:30]

Male: Just from a clinical trial perspective I feel like if you've got a chemotherapy control it's going to be difficult to use the new end points because you just can't assess them uniformly on the two groups. More and more we'll be in the situation where we have two immune based therapies that we're comparing and there to me it looks like we're pretty much ready to consider these end points and that they may

[05:24:00] be better reflective of patient benefit and more useful to scientific community, pretty much there.

Marc Theoret: Microphone 3, Suzanne.

Suzanne: This panel has focused on checkpoint blocking drugs and alternative end points and from somebody who has been a practicing oncologist, it seems to me that we

[05:24:30] didn't really focus on these issues until the advent of checkpoint blockers. In fact, in my past experience with cytokines and adoptive cell transfers and vaccines, I don't believe we saw these unusual response patterns, the mixed regressions or the so called pseudo progression. I'm wondering, this is to the entire panel, what is your opinion about whether the discussion today really is limited to considering checkpoint blocking drugs or do you think it's relevant to adoptive cell transfer and

[05:25:00] other forms of immunotherapy that we're developing currently.

Tony: Then you're right Suzanne.

Male: I agree with Tony, you're right. This is going to pertain to any biologic agent that's intended to lead to increased inflammation in the tumor. I guess the other side of it

[05:25:30] is any therapeutic modality that has a biologic rationale for having a delayed response and that includes a number of different modalities within the immunological field, probably some that fall outside of cancer and immunotherapy as well. I would guess that if therapeutic cancer vaccines ultimately are effective, probably in combination with checkpoint blockade, they should further enhance

[05:26:00] the type of pseudo progression and non-classical responses that we are currently observing.

Marc Theoret: We have time for two more questions. Microphone 2, Pat.

Pat: Yeah, thanks. I actually would like the panel to address having seen the data in the multiple different ways it's been cut, looking at how various approaches to immune PFS criteria, basically don't correlate particularly well with survival. If there are

[05:26:30] other thoughts that the group has about how to look at this, how to slice this, to try and find a better predictor of an intermediate end point that really correlates because it seems to me that we're not, we haven't gone as far as we were hoping to get. I'm wondering if in looking at the data if anyone has come up with some additional alternative approaches that really should be explored.

[05:27:00]
Male: Want to go?

Male: Sure, since nobody asked anything about my presentation. No, I think that's why my focus of my presentation is on survival because truly if we are not ... I think we focus a lot on the response rate, the tumor progression but if it is not capturing
[05:27:30] everything obviously we're missing something and that's something is something we don't know what they are and that's why we are trying to make the changes and try to figure out to see if we can capture that additional things. That's why we're having different cuts and trying to figure out if we can come up with a different criteria so that we can come up with a better correlation so that it can reasonably, likely to predict the overall survival.

That's why the focus of my presentations really ... also mention this briefly in the next section as well is that I understand Dr. [Broomenthal 05:27:59] from the FDA
[05:28:00] so you know conducted an interim, not interim, meta-analysis, looking at the milestone survival and also the overall survival and they come up with moderate correlation, right. I understand moderate correlation might not be good enough but perhaps that is the best we can get.

Male: At a first shot, it does seem like we have at least a couple of trials where you're
[05:28:30] comparing immune therapies and so we can at least look to see did the IRPFS [inaudible 05:28:31] IR Resist, hazard ratios correlate better with OS. That'll still be a small end at this point but it's a start anyhow.

Male: To address Pat's question directly, we actually did look at immune modified PFS in a single randomized trial, that's the popular study in non [inaudible 05:28:55] lung cancer for [inaudible 05:28:56] and actually had a very high correlation for what we
[05:29:00] called hybrid immune modified PFS and OS in that study. The reason we didn't present it is it does have its own issues, right. We know we're only selectively following the [inaudible 05:29:14] arm for treatment beyond progression and that benefit beyond progression and allowing for extension of immune modified PFS in that arm versus the control arm. Now, if we did it in the control arm maybe it
[05:29:30] would look just like a teso maybe it wouldn't ,because it's chemotherapy and chemotherapy we don't see those kinds of responses. We don't know. That's why ultimately we decided not to present that data but rather present what does it actually happen when you have a difference in immuno modified PFS versus standard resist PFS and show that correlation with survival. You are absolutely right. There's no definitive answer yet in any of these data sets.

Male: Can I add a comment on this?

Marc Theoret: Sure.

[05:30:00]
Male: Even at the risk we might go over by two minutes.

Section 33 of 45 [05:20:00 - 05:30:04]

Section 34 of 45 [05:30:00 - 05:40:04] *(NOTE: speaker names may be different in each section)*

Speaker 1: Even at the risk that we might go over by two minutes, so that question is a good question. I've asked myself that question every time we do it. It's ten years now. There's more data. We're getting more sophisticated, but we're not going to solve this 100%. A one to one correlation that we would love to have doesn't exist. It's perfect imperfection to say that. Having looked at it that way, I think we are at the

[05:30:30] PFS endpoint, as we're currently using it, the standard PFS endpoint is less perfect than the IRPFS, or IR resist, or IM resist, in terms of correlation with survival. There's actually more correlation with survival for the modified version than for the original, based on all of the data sets that have been presented.

Today, I think the best data set that we've seen was the Merck data set, but there are others that I know of, including the IM resist data that wasn't shown to the full

[05:31:00] degree today, so when you go through the whole thing, and we actually did a workshop on this in 2014, which was broader. There were more companies there that actually showed whatever they had available. It was very directional. Everything shows the same direction. There's a class effect here. There is a correlation of these new patterns with survival. The picture is more complete for

[05:31:30] that set of data than it is for PFS, or regular response, alone. It's not perfect, though. We realize that. The question is, then, where is the line? If there is a line that we need to cross for this to be more useful, where is that line?

Speaker 2: I just want to, before the last question, as Dan had mentioned, it sounds like there's a lot of opportunity here for collaboration, even a willingness across various commercial sponsors, which I think is much needed. That's the power of the FDA

[05:32:00] analysis, is really having that patient number to be able to explore, have these training sets, and testing. I think, your mention of the Cancer Immunotherapy Consortium is an excellent one, and a need to move forward with these analyses.

My last question before the break.

Speaker 3: Would anyone like to speak to what kind of evidence may need to be generated to better justify insurance coverage of treatment with immuno-oncology agents

[05:32:30] beyond the traditional definition of progression? Maybe that's something to think about over coffee.

Speaker 4: I think that's something that's come up in other meetings I've been to, especially when you go outside the US, where a country decides, "Well, this is what it costs to do this thing, and this [inaudible 05:32:47] of resources that we have. Can we do it or not?" The same way at my home, we decide can we buy a Lamborghini or not,

[05:33:00] and we answer no, we cannot. In places where there's a payer who decides, it makes sense to ask this question. If we start curing people, it's hard to put a price on that. If we do less than that, then we have a large number of patients who get a benefit for period of time, and then we have to decide as society, how much does that cost? That's not a regulatory question in the US, but it is a regulatory question in other countries. I know my colleagues from industry think about that quite a bit.

[05:33:30]
Speaker 5: I would totally agree with you there, Tony. The reason we're all excited and passionate about this field, the reason that this isn't just another cancer immunotherapy, just another targeted therapy is we've seen evidence, and we believe that there can be a tail on the curve. If we can't show that there are tails on the curve, if we can't show there are patients alive five, six years later, with metastatic, or at least formally metastatic disease, then this therapy isn't different,
[05:34:00] doesn't really deserve more attention than any other approach we have in cancer. I think for those of us who work in this field, and even as it comes to payers, I think that we need, as a community, to be able to show that level of benefit.

Speaker 6: I do want to make two comments about that question. This is an excellent question. When it comes to payer, I think there's another layer of that dimension. We're talking about demonstrating value, not just efficacy or safety, right? When you put a price tag on the drug, maybe a cure rate of 20% is not enough. If you
[05:34:30] want to charge the drug this much, I'm going to save 60%, right? I think demonstrating value is very different from demonstrating efficacy or safety. That's number one. Number two, in my experience, the majority of the payer organizations, they are still focusing on median survival, and I think there is a lot of education that needs to be done, because they are still looking at median survival, and based on that, to make a determination of whether or not there is a value demonstrated. If you are demonstrating a cure rate, or long-term survival rate, and
[05:35:00] that information is not being incorporated, then it's not going to be very helpful.

Speaker 7: I think the simple answer I would have to that would be the capital mar curves I had with the three components. That group that had non-PD by IR resist did much better than those who had progressed by both criteria. If that's not a health economic benefit, I don't know what is.

[05:35:30]
Marc Theoret: I think excellent discussion here, and definitely some next steps to be determined, but I want to thank all the presenters and panelists for this excellent session. Just for everyone to know, we're going to take a ten minute break, instead of fifteen, so we stay on schedule.

[05:39:30] [music 00:05:45-00:10:04]

Section 34 of 45 [05:30:00 - 05:40:04]
Section 35 of 45 [05:40:00 - 05:50:04] *(NOTE: speaker names may be different in each section)*

[05:46:30]
Sridhara: Get seated. We can start, and if we finish early, of course, we can all go home early too.

[05:47:00] If someone can call everybody in who is standing outside. Again, I think you all saw
[05:47:30] me, or if you didn't see me in the morning, I'm Raji Sridhara from FDA. I'm the division director of biometrics, which covers all of oncology, hematology products. Now, this last session is the most important session in looking forward to what we

can do from what we have learned until now in this workshop. This is for future
[05:48:00] clinical trials, how should we be designing them, and what are the considerations, and should we consider any novel designs?

In this regard, we'll start off with Dr. Ed Korn from NCI. Thank you.

[05:48:30]
Ed Korn: Thank you, Raji. Maybe I'll talk slow, as people come in. I'm going to talk a few minutes about phase three design considerations for agents with a potential predictive biomarker. There's my disclosure, it's that I have none. To set the context, we know that some agents mainly benefit a subgroup in the histologically defined population, and successful evaluation requires co-development of the
[05:49:00] biomarker to identify these sensitive subpopulations. There's various design strategies that you might use to integrate the treatment and the biomarker, and the choice of phase three design should depend on the biomarker's pre-trial credentials, that you think can really, can identify the subset, and that's what I'm going to be talking about.

I'm mostly going to focus on the case where we have a binary biomarker that separates the population into positive and negatives, where you think it's going to work and not work, and I'm going to assume that the analytic validity of the
[05:49:30] biomarker assay has already been established, so we know it's reproducible, and we know it measures something. I'm also going to assume that the biomarker credentials are sufficient that we can assume that the biomarker negative patients will benefit only if the biomarker positive patients benefit.

If you have a biomarker with very strong credentials, meaning that there's convincing pre-trial evidence that the benefits of the treatment, if any, would be
[05:50:00] restricted to the biomarker positive subgroup, then you use what's called an enrichment design, where only.

Section 35 of 45 [05:40:00 - 05:50:04]

Section 36 of 45 [05:50:00 - 06:00:04] *(NOTE: speaker names may be different in each section)*

Ed Korn: Enrichment design where only the bio-marker positive patients are randomized and the bio-marker negative patients are off studying. Examples of enrichment designs are the BRIM-3 study which was an anti-BRF for melanoma with ... Patient had a BRF mutation was how it's enriched and the KENO-10 trial which was in non-small
[05:50:30] cell lung cancer which is Pembrolizumab versus [inaudible 05:50:31] Paxil and there they enriched by PDL-1 expression being at least 1%. Only patients with PDL-1 expression greater than 1%, greater than or equal to 1% were enrolled on that trial. I'm going to return to that trial a little bit later.

The obvious limitations to the enrichment design is that since you're only treating the bio-market positive patients, you don't know if the treatment benefit extends
[05:51:00] to the bio-marker negative patients. You also don't know whether the bio-marker assay is doing anything for you at all, so whether it's the cost or inconvenience of using it pay off.

Before moving on, I just want to mention what I thought was an interesting design which kind of used to modify an enrichment design. This was CLGB30801 which was the trial of chemo plus or minus cell Cox IV in non-small cell lung cancer. The bio-marker was COX-2 expression and it was thought that it was going to work
[05:51:30] potentially, most likely work for the patients with COX-2 expression index greater than or equal to 4. They actually only enriched by COX-2 expression greater than or equal to 2, so people can be more randomized were 2 or more, but the primary analysis was 4 or more because they felt that under 2, they were sure that it wasn't going to work. This trial recently was reported, unfortunately came out negative, but it was an interesting design.

[05:52:00] If you have a bio-marker with strong credentials, but not very strong credentials, so you're not convinced that it won't work in a bio-marker negative population, but you're convinced that it would be at least as effective in the bio-marker positive population, as the bio-marker negative population. It might have some efficacy in the bio-marker negative population. Then you use typically a bio-marker stratify design or randomized all design. Here you measure the bio-marker first, and then if
[05:52:30] you're positive or negative you're randomized to the new drug or the control treatment.

It's kind of another way of drawing the schema which shows you randomizing the patients first into the new drug or the control, and then after that stratify, each one of those by whether or not they're positive or negative in the bio-marker and doing the analysis stratified by bio-marker status. It's essentially the same thing, the difference is whether you're going to require getting the bio-marker assessment
[05:53:00] before the randomization or are you willing to get it after the randomization. The benefit of getting it before the randomization is, I'm told, that encourages people to get it. The benefit of getting it after the randomization is you don't have to wait for it to do the randomization.

The ideal goals of a bio-marker stratified Phase III trial are to assess the benefit of the treatment versus control in each bio-marker sub-group to recommend the drug to patients who benefit and to not recommend it to patients who do not benefit.

[05:53:30] Now there's different formal analysis strategies for bio-marker stratified trial design, and that's what I'm going to talk about for a few minutes. A sub-group's specific strategies where you look at the effect of the treatment in the bio-marker positive group and the bio-marker negative group separately. It's something called the bio-marker overall strategy where you look in the positive group and you look in the overall group. For both of those strategies is a parallel version where you look at the two groups once and you split your Alpha, or you look first in the bio-
[05:54:00] marker positive group and then in the overall strategy group. I'll describe that in a second, and this mass design which is kind of a hybrid between the two.

In a sub-group specific parallel strategy you test the treatment in the bio-marker positives and the bio-marker negatives and you split the Alpha, so you can see here

[05:54:30] in Alpha I and Alpha minus Alpha I. For instance, using Bonferroni and ... The slides gotten a little mixed up with the yeses and no's here, but you get the idea based on how these tests come out, hypothesis test, you can recommend the treatment to the different groups either bio-marker positive or negative or both.

[05:55:00] In the sequential strategy, you test the bio-marker positive patients first and only if that's significant, then you test the bio-marker negative patients down here. The advantage of this strategy is you get to use the same Alpha here and here, so this has more power than the other strategy. You're still essentially testing the bio-marker positives and negatives.

[05:55:30] An example of the trial that used the sub-group specific sequential strategy was prime study which was chemotherapy plus or minus Panitumumab in metastatic colorectal cancer. The bio-marker with K Rath status, and so it tested the ... Since it was using the sequential strategy, they tested the bio-marker positive group first which is K Rath's wild type. This is PFS hazard ratio and the confidence [inaudible 05:55:40] for it. It worked in the positive patients, but it did not work and even went in the wrong direction in the K Rath's mutant patients.

[05:56:00] Now the bio-marker positive overall strategies, so here's the parallel version, you test the overall population Alpha and you test the positive population Alpha minus Alpha I, I'm sorry, Alpha I and Alpha minus Alpha I, what's left over, again, you could split them in half using Bonferroni. If you succeed in rejecting the all hypothesis for the overall population, you recommend the treatment for all patients.

[05:56:30] Here's the sequential version of it. You test the bio-marker positive group first, and then if that rejects, then you can test the overall population. This should say overall population here and you ... This would also be at the significance level Alpha.

[05:57:00] The KENO-10 trial which I mentioned a moment ago actually used a bio-marker positive overall parallel strategy on overall survival. It was actually a complicated analysis design which we're not going to talk about all the details here, but for overall survival, they looked at the bio-marker positive group and also the overall population group; and you could see for [OS 05:57:08], the hazard ratios were highly significant as we know the drug works.

[05:57:30] Here's an example of a trial. It was a trial of Letrozole plus or minus Lapatinib in metastatic breast cancer. Lapatinib being an anti-HER2 drug among other things, and for the HER2 positive patients they tested them first because they were using a sequential strategy. They showed a PFS difference and when this was significance, they tested the overall population and also saw a significant difference. The formal conclusion from this trial would be to recommend the drug for everybody.

[05:58:00] That's basically the problem with the bio-marker positive overall strategies. They could formally recommend the treatment for everybody including the bio-marker negative patients even though the treatment is ineffective in these patients. The

reason that's true is because even if there's no benefit in the bio-marker negative patients, if there's a very large effect in the bio-marker positive patients, it's going to make the overall group, the whole population, look like there's an effect. If it's large enough in the bio-marker positive groups, it's going to be statistically significant in the overall group, you're going to end up recommending the treatment for everyone.

[05:58:30] If we go back to the [patented 05:58:30] example, now the formally the design would stop after they said yes it works in the HER2 positive group and it works in the overall population group. The investigators rightly went ahead and looked into the HER2 negative group and low and behold it really didn't work there very well. They rightfully concluded that this drug should just be recommended for HER2 positive patients. Even though the formal design said recommended for everyone,
[05:59:00] they correctly overruled the formal design.

For the KENO-10 example, again the investigators, rightly in my mind, looked even though they didn't have to by the design, looked in the bio-marker negative population which in this case would be PDL-1 expression between 1% and 50% because the design was enriched for 1%, greater than or equal to 1%. They saw, actually here, it did work in the bio-marker negative population, so they could
[05:59:30] rightfully recommend the agent for [inaudible 05:59:28] population. With all the MERCK people here, is guess this is a good thing, right?

To get around this problem of recommending the drug for bio-marker negative patients where it doesn't work, we have something called the mass design, which first tests the bio-marker positive patients. If that rejects, if it works for them, then
[06:00:00] you go ahead and test the bio-marker negative patients. If that doesn't work, then you test the overall population. This seems ...

Section 36 of 45 [05:50:00 - 06:00:04]

Section 37 of 45 [06:00:00 - 06:10:04] *(NOTE: speaker names may be different in each section)*

Ed Korn: Doesn't work, and then you test your overall population. This seems a little bit screwy at first, but if you think about it, even if you don't reject the bi-marker positive patients, as the treatment working, it may actually work there, just not well enough; and then you go to the overall population. If it works there also, then together you'll get a significant result in the overall population and you can recommend the drug for everyone.

The mass design, I went through that quickly, this references at the end, it
[06:00:30] minimizes the probability of recommending ineffective treatment for the bi-marker negative patients, but it maximizes powerful treatments when the treatment effect is similar in the bi-marker positive and bi-marker negative sub-groups. An example using mass design with an ECOG Trial in [ALL 06:00:43], the bi-marker with minimal residual disease, the trial is on-going.

Kind of the take home message here is that you should, obviously, select your design based on the bi-markers credentials before the trial begins. If you're going

[06:01:00] to use a bi-marker stratify design, you really need to make sure that if you're going to recommend the treatment for the bi-marker negative patients, you really need to make sure you don't want to be doing that if it doesn't even work there.

Thank you.

[06:01:30]
Sridhara: Next speaker is Dr. Lillian Siu and she's from Princess Margaret Cancer Center in [inaudible 06:01:34], Toronto.

Lillian Siu: Thank you very much. I want to thank the organizers for the opportunity. I wish I had a crystal ball for this session, but I don't; and having learned quite a bit from the last few days, hopefully, I can share some of my thoughts.

My topic is on IO combinations in terms of future and near-future clinical trial design considerations.

[06:02:00] Oops, going backwards. Hopefully that's not going to be the case for the talk.

These are my financial disclosures. I think over the last two days we've heard a lot about the complexity of the adaptive immune system and that already we have many of these targets that have drugs for. The challenge lying ahead of us is really how to combine some of these drugs together or together with other agents such as targeted therapy, chemotherapy, or even radiation.

[06:02:30] Borrowed from Pamela [Hashi 06:02:29] from my institution as a medical oncologist, my simplistic thoughts are we have immunogenic tumors, that are so-called hot tumors, that perhaps we already know a certain proportion of patients will respond to immune checkpoint inhibitors, ACT, etc. The bigger challenge is really the larger group of patients who have the so-called non-immunogenic, or cold tumors, and how do we make these tumors more green, more immunogenic. Some ideas already have been discussed in this meeting. For example, using
[06:03:00] multiple therapies combined together to prime the immune system, activate it, and then subsequently [inaudible 06:03:05]. For example, follow with immune checkpoint blockade.

For the topic today that I'm going to discuss, I'm going to focus mainly on IO based combinations, on how potentially to do so. Listed on this slide, I'm sure is not exhaustive, are really just some of the drugs and cell therapies and different
[06:03:30] treatments that are being combined with PD-1 PD-L1 combinations, in combination. I'm not a bio-statistician, but I can easily do the bio-statistics in math, that if we are to even look at these in doublets, we are talking about a large number of potential clinical trials. If you start looking about triplets which is really already starting to enter the clinic, or at least being planned to go into the clinic, that number becomes even more prohibitive.

I think the challenges with combination is 1) how to select the right combination to

[06:04:00] move forward; and if you're able to do so, how to actually do those trials efficiently if the clinical trial design can help you. Lastly, how do you have the readout that you know your combination is better than, for example, PD-1 PD-L1 alone. I think those three points are really going to be important for us to move forward to look at IO based combinations.

This is from Twitter. Just to show that there are a lot companies, a lot of drugs, and some work together, some don't work together, some work together with
[06:04:30] themselves in trying to combine their agents together. I think they all have the same questions. How do I know that two or three of these drugs together is better than one of them alone? Obviously, there is another side of the equation which is the therapeutic index, the toxicity. As you increase the efficacy, same potentially can go with toxicity, and we need to keep that therapeutic index in mind.

First of all, the rationale, why would you take a combination forward? In the
[06:05:00] chemotherapy and targeted therapy era, we would think of rationale such as synergistic effects, synthetic lithology, and obviously, reversal of either a primary or adaptive resistance.

I listed some examples here in the targeted and chemotherapy era. What would we do would consider to take something forward as synergistic? For example, Dual HER2 blockade in breast cancer, B Rath neck inhibitions in melanoma. Those are classic examples that already have been tested and validated in the clinic. Perhaps
[06:05:30] in the IO case, combining a neck inhibitor and immune checkpoint blockade, we're seeing some promising, potentially synergistic effects that would give good rationale to move forward.

Synthetic lithology, such as part inhibition, plus RT or DNA damaging agents, I had a hard time trying to figure out what is synthetically lethal in the immune system [inaudible 06:05:48]. The only thing I can think of, potentially, is TGF Beta inducing a [inaudible 06:05:53] resulting in a synthetic lithology if we combine that with part
[06:06:00] inhibition. I think this needs still to be validated and credentialed.

The reversal of resistance we all know that cell cycle inhibition and ER inhibition in breast cancer is one potential such mechanism and perhaps in the IO scene, the combination of an anti-[inaudible 06:06:11] antibody with and anti-PD-1 PD-L1 inhibition would be such a mechanism.

I think the point of this thought is really, we need to think about the rationale and not just pick two or three drugs in our pipeline that we can get together as being the reason to move two drugs forward, or three drugs forward.

[06:06:30] This slide is summarized from the various data presented at ASCO, ASH, and other meetings in 2016 and obviously not complete. Even just looking at Pembrolizumab alone in combination of various other agents, either in doublet or in triplet, where we are seeing very promising single-arm response rate. Obviously, these kinds of trials are only going to increase in numbers over time. We don't have all the

[06:07:00] patients in the world to do every trial. We don't have all the money in the world. How do we fine-tune and how do we streamline to take the most interesting combinations forward?

I'm going to talk a little bit about the toxicity aspect that I've highlighted earlier. This is the study that I think several presenters have actually shown in this particular meeting. The clinical trial CheckMate 067 Melanoma where single-agent [Nivo 06:07:25] versus single-agent [Ipi 06:07:27] versus doublet [Nivo 06:07:28]
[06:07:30] [Ipi 06:07:29], and you can see the various Kaplan-Meier curves here.

I'm going to focus more on the toxicity data that was already presented and published. If you look at most of the immune-related and treatment-related adverse events, if you look at the simple mathematics, most of these toxicities are additive, so that column, plus that column, equals this column. We do see that, for example, in the case of hepatic toxicity, perhaps we're seeing a little bit more than
[06:08:00] just additive. I think these are things we do need to keep in mind as we take combinations forward, not just the efficacy aspect but, clearly, also the toxicity aspect we need to think about.

Another example where we combine Ipilimumab with a B Rath inhibitor [inaudible 06:08:15], and where we do see even in a small number of patients increased liver toxicity. This is from Tony's paper in NEJM a few years ago. Really just to bring the
[06:08:30] thought out there that we sometimes forget that toxicity is as relevant as efficacy as we combine these drugs in combinations in the clinic.

I pulled a few examples of Phase I trials that have already been either published or presented in abstract format and look at what they did use in terms of Phase I dose finding design of these IO based combination. This is the paper of the first Phase I study in melanoma that looked at [Ipi 06:08:58] and [Nivo 06:08:58] together. In
[06:09:00] this particular case, three plus three was used initially, but then it was changed to allow cohort expansion as time went on. As we heard yesterday, both agents underwent dose escalation, but the dose of [Ipi 06:09:12] was already tested as a single agent. It was the [Nivo 06:09:14] that was not yet tested. This was really careful in terms of how to test the different doublets and different schedules, both concurrent and sequential together.

[06:09:30] In ASCO this year, there were at least two of these IO plus IO combination doublets that were presented. One is a 41BB Agonist plus Pembrolizumab presented by Tony [Tocher 06:09:39]. Twenty-three patients and solid tumors, patients with prior immune checkpoint inhibitors were allowed. They used a tight CRM design, so this is a Basion model based design that really allows patients to go on and dose escalate even after the first cycle; but in totality, you would be able to look at the
[06:10:00] more delayed toxicity that happens during the trial. In this particular case ...

Section 37 of 45 [06:00:00 - 06:10:04]

Section 38 of 45 [06:10:00 - 06:20:04] *(NOTE: speaker names may be different in each section)*

Lillian Siu: It was during the trial and in this particular case, the Pembro dose was fixed and

the single agent PF dose was already determined in terms of what is the recommended Phase 2 dose, and in this study, the dose being escalated is the 41BB agonist.

And then in other study, which is a study we participated on with an OX40 agonist together with atezolizumab, an NTPDL1 inhibitor, twenty-eight patients saw their
[06:10:30] tumors, again patients with prior ICI immune checkpoint inhibitors were allowed, again used a three plus three design, after the single agent OX40 anti-agonist has been finished and then in this particular combination study, the atezolizumab dose was fixed and then the OX40 agonist dose was escalated.

The point obviously is that there's no right answer. We heard yesterday that people don't seem to like three plus three because you don't have a lot of patients per
[06:11:00] cohort to really understand the toxicity and the efficacy, that we wanted a little bit more than three patients. But again, you know, if you look at the literature, I don't think that the jury is in yet.

With that, i think the next few slides I'm trying to figure out how best we can gather that evidence to create that knowledge base, that helps us to answer the three points I have mentioned. Which is one, how to choose rational combinations to move forward. Two, how to do it efficiently. And three, if you do decide to choose that combination, how do you have a read out that tells you the
[06:11:30] combination is safe and efficacious.

So looking at different ways to do this, one is to look at the literature and do in silico analysis. There are published data on these immune modulatory targets and TCGA. Obviously we know that in TCGA, look at primarily tumors but many of these samples looked at mixed tumors and normal cells. Perhaps from that we can glean from in terms of IO targets that are expressed. Look at tumors that have different
[06:12:00] expression profiles of all these immune modulatory targets. Generate some heat maps. Look at it based on disease sites. Even based on mutational status, for example, of the disease sites. And see if we can learn something in terms of the pattern of these immune modulatory target expression. And we can obviously try to actually validate that the expression means something. For example, using that kind of data, comparing with the efficacy of PD1, PDL1 inhibitors out there.

We're actually taking on this kind of in silico analysis to try and build a database,
[06:12:30] built knowledge in terms of what is a rational IO combination to move forward.

The other thing is really to take samples that are already ongoing in clinical trials or design your own clinical trials to get that data. So we have an in house study called Inspire, which is a study with pembrolizumab, with the main end point not to look at efficacy. Mark has done a lot bigger studies than 100 patients to look at efficacy in these tumor types I have listed here: squamous cell cancer of the lung, triple
[06:13:00] negative breast cancer, epithelial ovarian cancer, malignant melanoma, a mixed tumor basket.

The goal of Inspire is to do deep dives. What we wanted to do and we are doing is to take biopsies at baseline at the first imaging or just before the first imaging and those who respond at the time of the progression. We will do sequencing, transcript [inaudible: 06:13:20], circulating DNA, radio mix, etc. To really understand at a molecular level what is happening. And more importantly, we hope
[06:13:30] that we can use this to set a base line. Because for you to know what the combination adds, you need to know what the base line shows. In fact, I put Tony on the spot during the break to say 'what is the amount of tills present that is really relevant and how much increase do you need to see in the tills after a combination that would be considered important?' I'm glad to know that I don't know the answer. Neither does he. So that's great.

So I think we need to build the knowledge base such that we know at the very least
[06:14:00] what do we expect and I understand totally that this may differ based on drugs, type of treatment, ACT may be different from a immune checkpoint, different from an agonist. But nevertheless without even this base line it would be hard to know how much a combination adds to the base line.

I think the third way to do this is eminence based medicine, which is we are actually doing a survey of people who have done a lot of these trials and say 'look, if you have to do combinations, what do you think is the most efficient
[06:14:30] mechanism?' So, with the investigational drugs during committee CTAB, we're actually planning a survey amongst drug developers, many of whom have used a lot of these combinations in the clinic, to really understand by their experience and their insight, what they think is really the most efficient dose escalation methods to take combinations forward.

One design that I like is the zone design for combinations where you have two drugs, where you have different doses of the two drugs, especially good if they are
[06:15:00] investigational to both begin with, so you have low dose, medium dose, high dose of one drug and exactly the same for the other. Where you would obviously start off with low, for example, for both and then as you escalate, you can go to two zone, medium for both, or even low on one and medium on the other and vice versa. And you would only branch out to the higher doses as you clear the previous combinations or previous doublets. And I think obviously time is of essence. If we
[06:15:30] do this kind of things sequentially, where you can only do one, two, three, four, five, six, seven, sequentially. Time would potentially be lost. But you do need to have some confidence that you can do [inaudible 06:15:42] open multiple cohorts at the same time, that you have the safety assurance that you could do so.

I'm close to my last slide and I have two minutes left. This is a review we just wrote recently for Genome Medicine, really thinking about how to take combinations
[06:16:00] forward. I don't think I have enough time to really project what will happen in the future. I think some sort of adaptive assessment of multiple schedules or combinations is going to be important. Once you have the ability to actually take what is interesting forward to the clinic and use the efficient design to do so. And I think the important thing is really drop the choice of the regiment that is not really

moving, looking interesting quickly and move the optimal combination using an adaptive design to study it well in different populations who are likely going to
[06:16:30] benefit.

I think it is a difficult area to look at combinations, particularly difficult when you start moving into three drugs or four drugs. But I think we need to start thinking about doing this in a more systematic way such that we can gather all that data and evidence to put it and learn together.

So to conclude, I think IO based combinations currently are really largely based on
[06:17:00] empiricism. Rational designs, for that to occur, I think we do need to build this knowledge base as a base line and we all heard, although some new non-clinical models are being created. The [inaudible 06:17:07] on non-clinical models are by and large not ideal. I think clinical trial considerations are important, not just for the efficacy but we need to think about additive toxicity and how do we have a short term read out that is reliable. Is it really the clinical parameters that we just heard in the wonderful session just now. Is it some sort of immuno-phenotyping and to do so we need to know what a base line immuno-phenotyping profile would
[06:17:30] look like. And as I mentioned, I think the era of triple combination is coming and the methodology, I think, is lagging behind. And we need to really be able to build a knowledge base to catch up.

Thank you for your attention.

Sridhara: The next presentation is from Doctor John Kirkwood from the University of Pittsburgh.

[06:18:00]

John Kirkwood: Well thanks for the invitation to speak at this wonderful two day symposium. I have to say statisticians keep us honest and keep us on our toes. And we've had a ample dose of that kind of inspiration here over the past two days. My disclosures and basically where we start in melanoma is in metastatic disease but this is a very
[06:18:30] small part of the population. The area that we spent the last, sad to say 30 years, is trying to develop therapies that are going to work in the adjuvant arena so I'd like to spend half of my time on metastatic disease and how we might do that a little better. And the adjuvant conundrums which have, even from last Saturday, begun to make the air a little bit thicker is the last half of what I'd like to talk about.

You already heard about the many agents I'm forced to muse that Ed [inaudible
[06:19:00] 06:18:55] and I gathered seven statisticians and seven leaders of the cooperative groups to assemble thirty-five years of data to say, not one agent that we had tested prolonged survival or improved progression pre-survival, only in 2008 and today we have 10 agents, soon a dozen, approved in the metastatic space to different classes of agents in the adjuvant space and we don't really think about chemotherapy much anymore. The problem is that for advanced melanoma we
[06:19:30] don't have the time to assess as Lillian has already just introduced. We can't cope with the number of single, never mind binary and ternary combinations. The time

in metastatic disease on average for new approvals, seven years. Patients, hundred per arm, resources we really need to get around this obstacle.

In the ECOG-ACRIN melanoma committee that I've had the pleasure to chair for the last thirty odd years, basically we have a ternary combination already in trial. We
[06:20:00] have actually two doublets I should say in trial. The ...

Section 38 of 45 [06:10:00 - 06:20:04]

Section 39 of 45 [06:20:00 - 06:30:04] *(NOTE: speaker names may be different in each section)*

Speaker 1: We have actually two doublets I should say in trial. The trial that Mike Atkins is currently leading which is an intergroup effort of SWOG and ECOG and all the other NCI funded groups. As a landmark, two year, overall survival end point, asking whether the doublets of molecular therapy for BRAF and MEK, or the doublet [Ippy Neevo 06:20:23] that we've heard about so much in this last two days are better
[06:20:30] for patients up front. This trial, schematized here, randomized patients who are BRAF mutant up front to receive dream doublet DT or dream doublet IN and at progression they cross over so no one is cut out from one or the other. It's a question of which do we do first and this really is a fundamental problem confronting the whole community.

For the wild type melanomas is where we have our [ternary 06:20:53] combination
[06:21:00] based upon very interesting data Steve Hodi published JAMA Oncology a couple of years ago where Ipi high dose combined with GMCSF did something we really still haven't fully explained. It gave us a better overall survival outcome, compared to Ippy alone, when we didn't see any impact upon progression free survival, or any of other conventional end points in the study in metastatic disease.

[06:21:30] EA6041 currently ongoing, and other high priority study of the cooperative groups takes Ipi Nivo with GM with Ipi Nivo alone and ask if we see this again in a second large trial. This will focus upon an initial cohort of 240 patients, where we are already booming well along toward this answer if positive at this interim analysis stand point; which we hopefully will have within the middle of the next year, go to 400 to a formal phase 3 study, but this is one way to be a little bit more efficient.

[06:22:00] In the international melanoma working group, which we have happily been able to gather with every 6 months for the last 11 years. A trial called the melanoma international collaborations were adaptive trails, MICAT, has been something we have been thinking to bring forward for a while. This lead by Grant McAuthor, Mark Middleton, Derk Wrightsmith from PPD, Don Berry from Berry Biostatistics and
[06:22:30] both patient advocacy, with Valery [Gilled 06:22:27], and RA Capital is actually confronting the problem I just mentioned. That we have no capacity to deal with the combination therapies and we need to do things more efficiently in smaller numbers for the future to avoid the problem with failed phase 3 studies, which is still unfortunately fairly common.

Adaptive trial approach enabling multiple combinations based upon state of the art bio markers and option for more rapid regulatory approval, as Janet Woodcock has

[06:23:00] published with Eye Spy in the literature, involves all of the institutions illustrated here from US, Canada, Australia, Europe. Basically is a collaboration of 18 academic institutions, as I said already, advocacy biostatistics industry, which we would like to see more involvement from PPD as a CRO and input from the FDA and RA capital
[06:23:30] to develop a collaborative phase 2 study that would then potentially evolve to phase 3 concurrently evaluating multiple candidates with a single common control arm. Would assign patients to each sub population on the bases of the highest probability of success as Eye Spy has already done. Test treatment interactions drop, or graduate, new treatments according to their benefits and address potential sequencing down the road. The goal of this to speed therapies to patients. The bottom line to have adaptive statistical methodology, which is more
[06:24:00] efficient, will require fewer patients per arms then our hundreds patients in metastatic disease, currently required to answer questions with a higher success rate projected down the road. Where the cost to sponsors to industry would be less.

The alternative, I think we all know. Cooperative groups, where we continue to work hard; industry supported trails, which are obviously another way to go; IITs,
[06:24:30] as Lilian just mentioned, but this design to take patients in and to say that principle problem we're faced with is what to do post PD1. We won't tackle front that [Denovo 06:24:36] untreated patients, we take patients failing PD1, adaptively randomized A, B, or C and those A, B, or C are illustrated here in a figure that Julie Kieth has put up, and which you can find on micat@ppd.com if you would be interested to look further, where we cope with both BRAF mutant population, wild
[06:25:00] type population, can consider vaccines as a possible options and would use a genomic and cellular bio markers throughout.

This requires a bit more funding, one or two more participants from industry, but it certainly has precedence, more recently GBM agile, that I think encourage us to continue in this direction.

The second part I wanted to talk about is the place we spent our last 30 odd years
[06:25:30] developing, or trying to develop, more effective [adjuvant 06:25:31] therapies. Thinking about the thousands of patients that have been on the trials that have lead us to two different classes of regulatory approvable agents with millions of dollars spent on this and the time, a decade, between the time we throw the stone and the time we hear the splash. I think this is the area where the impact, potentially much huger, it is even more confounding.

[06:26:00] Many questions we have talked about earlier and I think, given time and the late hour of the day I won't spend a whole long time, but say that we have the principle hypothesis for all of these adjuvant trails that the impact in the adjuvant arena will be a greater one then in the metastatic disease setting that we are currently testing [Pembro 06:26:19] verses alternatively approved agents interferon, or Ippy, [EORTC 06:26:25] is going to do a similar trial, BMS has a Neevo verses Ippy trail. The
[06:26:30] [comby 06:26:30] add and the brim 8 trial with molecule targeted therapies are out there and pending as well.

Does interferon work better in the adjuvant setting than in the metastatic disease setting? It's never been approved in metastatic disease, so it's a hard thing to test; with high dose Ippy %11 response rate and had ratio that you all have seen several times. The questions of whether the benefits in adjuvant setting, reported only last Saturday in the [Matura 06:26:56] trial at ESMO are remarkable positive and maybe
[06:27:00] more positive for the OS than even for the RFS, as you see in the bottom of this trial, in a now mature data set out passed 5 year of median fallow up.

These two approaches, high dose interferon Ippy, have remarkable similar benefits to patients and I think the data from Ippy trial is worth looking at because this is
[06:27:30] high dose Ippy 10 milligrams per kilogram in stage 3 patients. The data we saw several years ago at ASCO for the RFS impact, which was the primary goal of this trail, followed now by this data presented for the RFS at maturity very similar and about %11 improvement at 5 years. This data at 5 year follow up for survival, again about a %10-11 percent improvement at 5 years, which is remarkable similar to the
[06:28:00] data that was the pivotal trial, now almost 20 years ago, with high dose interferon.

Deaths on Ippy, an obstacle to consider what to do with, and the question really then, bears down upon the question that confronted the regulatory authorities in the US. Is 3 or is 10 a better dose? This trial, which is still in the hopper, is actually fully accrued 2 years ago, is our US cooperative group, inter group I should say, trail complete with accrual 2 year ago and about another year and a half before we
[06:28:30] think we'll see OS data because, with Ed, we decided OS was the end point to look at in trial. A code primary very small alpha devoted to RFS and this data [Pala Acherto 06:28:43] presented just last weekend in Copenhagen. Basically bear upon this because this is a very large trial conducted by BMS of Ippy, testing 10 verses 3
[06:29:00] in the metastatic measurable disease setting [Denovo 06:29:01].

These data, I think we will all have to all chew on for a bit, but clear superiority in terms of survival for the higher dose Ippy dosage. This data ,that basically we'll have to think of in terms of metastatic disease, may or may not, have impact in the adjuvant disease setting. No forest plot differences to make any big deal about, but
[06:29:30] PFS overall response, durable response disease control ... No differences that are significant. The only impact of this trial appears to be in overall survival. The expected differences between high dose Ippy and low dose Ippy, in terms of treatment related [AEs 06:29:47] is substantially more for the high dose Ippy. I think the question of which dose in metastatic disease we ultimately come to think about in the US, where we have only 3 mixed per kilo approved is something we
[06:30:00] have to consider, also we have to consider what do with this in terms of the [inaudible 06:30:03]

Section 39 of 45 [06:20:00 - 06:30:04]

Section 40 of 45 [06:30:00 - 06:40:04] *(NOTE: speaker names may be different in each section)*

John Kirkwood: -have to consider. Also, we have to consider what to do with this, in terms of the

adjuvant arena, where only 10 per kg of Ipi is available. How to address this more efficiently? I think the answer to this, for the last 15 years at our center in Pittsburgh, has been to do things in a neo-adjuvant setting. Take patients with bulky, regional measurable, operable disease, and instead of operating them first,

[06:30:30] treat for a month or two. The impact of this, as you know, in many other tumors, is clear in terms of survival, respectability, local control. Certainly can accelerate progress, and experimentally, it finally gives us tissue, both before and after the intervention.

We did this first with high dose [interferon 06:30:45], the data before and after for T cells for dendritic cells requires no description. For stat 3 significantly less. We've done it and published it for high does Ipi, 10 per kg in neo-adjuvant trial 08144,

[06:31:00] here, even at 6 weeks, major responses, so maybe earlier disease can respond even in this neo-adjuvant setting, meaningfully.

The trial to mention that bears upon this question of whether 3 or 10 of Ipi is the dose to choose, is the neo-adjuvant trial that [Amad Tarheeni 06:31:16] has run and presented at [ASCO 06:31:18] this year, 11063, which took the exact same design, biopsied before, biopsied after, and had half of people assigned to 3 per kg,

[06:31:30] half to 10, given with high dose interferon. The distribution of patients is illustrated here, but since this was at ASCO, we're running out of time, I'm going to skip on.

The grade 3, 4 toxicities with 3 per kg - substantially less than for 10 per kg of Ipi, no surprise exactly as the literature has for us, the data. In terms of radiologic response, the differences between 3 per kg and 10 per kg of [Ipilimimab 06:31:58]

[06:32:00] here, slightly better for the 10 per kg, but not bad for the 3, and for the pathologic complete response, which of course you can look at in the neo-adjuvant setting, identical - or even better for the 3 per kg.

I would submit pathologic complete response may well be an endpoint for neo-adjuvant trials that we adopt in the cooperative group in the future. I think neo-adjuvant therapy is really the way we need to accelerate progress in the adjuvant arena.

[06:32:30] The current trial, [penbero 06:32:30] versus high dose, led by Tony and his group as [SWOG 06:32:35] has a primary endpoint of overall survival. RFS QRL are included, and to build more effective adjuvant therapy, myriad combination that I could have talked about, but I think time probably doesn't permit, so I'm going to skip through all of this and say it'll be there in the set of slides that you have otherwise. I think neo-adjuvant approaches are the only way we can cope with the dissection of the

[06:33:00] multiple combinations that lie before us. I think that these are the way that we need to go for the future to dissect all of those possibilities that Lillian, and others before me, have gone through throughout this two days. Thank you very much.

[06:33:30]

Sridhara: We have our last presentation by Dr. David Rimm and this is a topic that we have not heard today and yesterday on some of the companion diagnostic tests. Thank

you.

[06:34:00] Okay, I think I was racing. I missed one presentation in between. Tai Chin is going to present from BMS. Yeah, thanks.

How did I do that?

Tai-Tsang Chen: Tai-Tsang Chen from BMS.

So the topic of my presentation is designing late stage randomized clinical trials with the cancer immunotherapy. Can we make it simple? Listening to a lot of the clinical challenges and difficulties that we just heard from previous presenters, you can imagine how difficult it's going to be for statisticians who designed all these trials.

[06:34:30] For background, I'm going to show you some of the hypothetical situations that we're observing over survival curve. This curve is exponential distribution, representing chemotherapy, and this is target therapy, I believe, that Dr [Areba 06:34:42] has also shown something similar as well. You'll see an early effect in the overall survival curve. This is the third curve. The blue curve represents the immuno-oncology agents' overall survival. You see a delayed [inaudible 06:34:56] effect at the beginning and then there's a plateau at the end.

[06:35:00] The yellow curve has been shown multiple times. This is the immuno-oncology PFS curve. There is a sharp drop at the beginning and then there is a plateau towards the end. Essentially, what we would like to see is that- Let's say we combined a targeted therapy and immuno-oncology agents, hopefully we're going to see some early effect and also there's a plateau at the end.

[06:35:30] Essentially, the problems that we're facing right now, is - Number one, there's a delayed [inaudible 06:35:26] effect, and number two there's a plateau, so how do we deal with those two unique characteristics? What are the lessons learned? Number one, as I mentioned, unique survival kinetics. Number one, long term survival. What have we been doing, or what have we been thinking about doing?

[06:36:00] In terms of study design, we want to make sure that we enrolled an adequate number of patients, so that we have enough patients at risk so we can get to the number of events that we are targeting. Those were mentioned earlier as well. Number two, we would like to potentially identify who are those long term survivals so that we can target those patients and give those patients drugs and basically move one step further to personalized medicine. Number three - that's what we're here for - we are exploring some novel end points.

What about the statistical analysis? What statistical analysis are we thinking to deal with long term survival effect? I mentioned earlier about [inaudible 06:36:20] survival, and also the restrict immune survival. Essentially, the life expectancy. Also, it has a ratio over time as well.

[06:36:30] What about the delayed clinical effect? What were we thinking about? In terms of study design, I think Dr. [inaudible 06:36:34] - who's also mentioned this as well, earlier - that we try to quote, unquote "overpower" the study so that we anticipate that delayed [inaudible 06:36:42] effect. We account for the delayed duration in a synthesized [inaudible 06:36:46], and we increase the information fraction at the interim analysis. For those who are not statisticians, information fraction represents a number of events at the time of interim analysis. If you move this time
[06:37:00] towards the end of the study, what happen is that you will accumulate more number events, so the number in us is going to be more robust.

What about the statistical analysis? Again, we could potentially look at the [inaudible 06:37:13] survival restrict immune survivals to a point that you're not looking at the time when they're still overlapping. You look at two years, three years, at the time when the curves have already separated. Again, you look at the [inaudible 06:37:25] ratio over time. This has been in many, many presentations
[06:37:30] and many workshops as well is that people are considering using weighted [inaudible 06:37:33] tests. Essentially, you are weighting more to patients towards the end of the curve - patients who live longer.

Obviously you are encountering philosophical questions, of whether or not you should treat patients differently, right? Why should we treat patients who live longer more than patients who live shorter? Finally, in order to prevent analyzing
[06:38:00] the data at an early time - when the curves are still overlapping - you want to ensure that we have sufficient follow up, so that we give the curves a chance to separate.

All these are things that we have learned. Lesson learned number two - the tail of the curve. The primary focus of the overall survival trial has shifted from the improvement in overall survival to raising the tail of the curve. I've heard about this again and again. In my opinion, there should exist an improvement in the tail of the
[06:38:30] curve that is considered clinically meaningful by treating physicians.

I'm hoping this is indeed the case. That if I talk to treating physicians then I said, "Okay, what is the improvement that you would like to see in terms of survival rate at two years, or three years?" There should be a hypothetical number in their head. The last point I want to make is that the tail needs to be accounted for somehow in the design or analysis. If the purpose, from this point on, is to raise the tail of the curve.

[06:39:00] What is the tail of the curve? I mentioned this earlier as well - it's a milestone survival. Milestone survival, again, I mentioned earlier is that it's defined as the capable [inaudible 06:39:07] survivability probability at a specific time point. Again, in these presentations I'm not going to answer a question of "What is that time point?". This is a very difficult question, it has to take into account clinical, statistical, factors.

[06:39:30] Is two years sufficient? I remember 6 years ago when we were developing [ipi-limo-map? 06:39:26] and we had several meetings with key opinion leaders. At the time, we believed two years is the right time, but if you look at anybody I think two years probably is too early. You may want to look at 3 years, for example, so it depends. In my presentation, I'm not going to answer the question of what would be the optimal time point for this type of analysis.

[06:40:00] Previously I've mentioned that this can be used as an intermediate endpoint at the interim analysis in a subset of patients. Why the entire study still use overall survival as a primary endpoint? What happened is that, as an example, you have 5 hundred patients in your study, you look at the first 3 hundred patients, at the-

Section 40 of 45 [06:30:00 - 06:40:04]

Section 41 of 45 [06:40:00 - 06:50:04] *(NOTE: speaker names may be different in each section)*

Tai-Tsang Chen: The patients in your study. You look at the first 300 patients at the interim analysis, and the reason why you look at the first 300 randomized patients is because those patients will have longer follow-up, right? If you look at all 500 patients in your survival interim analysis, what happens is that you will have patients, the last 200 patients that randomize through the studies will have shorter follow-up, so you will have an earlier censoring, for example. That would make your analysis less robust.

[06:40:30] The proposal is that you look at the first 300 patients, as an example, and you look at the survival, the milestone survival, and see if there is a hint of efficacy. You can actually build it into a design accounting for your Type I error rate so that there is a stopping rule built in as well. The pros are, it's a preview of what the tail may look like. That's what we want to see. It allows us to have an early access to efficacious treatment to patients, obviously. The cons are the totality of the data are not accounted for. I mentioned in a previously presentation as well that Dr. Brumenfeld [06:41:00] from the FDA conducted a meta-analysis in [inaudible 06:41:05] lung cancer, and they found moderate correlation between milestone survival and over-survival. Approximately somewhere between 60-70%, if I remember correctly, and censoring private milestone impacts survival estimate, right. That's why in my previous presentation I analyzed data at two years, and I make sure that all [06:41:30] patients have been followed for at least two years. If you look at this earlier, then you have a lot of censorings that would potentially skew your-- impact the robustness of your analysis.

[06:42:00] Can we make it simple? This is a very important question, at least to me personally, because with all the challenges, and we have to make a lot of assumptions, but can we make fewer assumptions in your study design. Because, after all, those assumptions are our hope, based on our educated guess, but they are guesses. What we would like to do, or what I would like to do personally, is that, can I design a study or come up with a study design that over-survival remains the primary end point, and it is less sensitive to the uncertainty of survival kinetics? We are talking about delayed term and effect, we're talking about long term survival effect, and that the tail of the curve is accounted for. This is the objective that we'd like to achieve.

[06:42:30] Here is your proposal. Can we design a study using milestone survival, analyzing the data with co-primary end point as the milestone survival and over-survival? In order to design any study, you need to have a couple factors, treatment effect, follow-up duration, false positive rate, and finally you will be able to estimate your sample size.

[06:43:00] Let's take a look at this. This is an illustrated example, treatment effect and follow-up duration. These are the first two factors that I just mentioned. Assuming that I talked to a physician, and we believe that a meaningful improvement at two years is from 31% to 51%. Let's say 31% is what we have right now, the best treatment effect we can achieve, based on the best treatment right now we have. I would like to see a 20%. That is the expectation. That's simple enough. Look at this. It can include all kinds of curves, all kinds of shears that you can have. If there's no treatment effect, that implies, for simplicity, I am using exponential distribution as an example. Then I am looking at a hazard ratio of post-delay. When there is no delay, that means that the separation starts at the very beginning. I am looking at a

[06:43:30] hazard ratio of .57. If there is an eight-month delay, I am looking at .27. Regardless of all these scenarios, what I would like to see is an improvement from 31% to 51%. Easy enough.

[06:44:00] What this tells you is that I have defined the treatment effect, the size of the delta. Number 2, I know what my starting [inaudible 06:43:52] is going to be. In order to have the most robust result, that means I need to follow the patient for at least two years. When I hit the two-year mark, then I stop the study and analyze the data. That is agnostic to the survival kinetics, as I mentioned. I don't care anything about the delay and also the long-term survival effect. As I mentioned, I really want to limit the number of assumptions I am making. The assumption I'm making here is 31% and 51%. I am not talking anything about the delayed duration or the cure rate of control versus experimenter.

[06:44:30] Number two, what about the false positive rate? Because these two end points are correlated, as I mentioned earlier, we have observed moderate correlation. Why don't we just take advantage of that? The plot actually shows the spectrums of the correlation going from zero to .9 on the X axis. On the Y axis, it goes from family-wise, or the Type I error rate, ranging from zero to .05. Let's say I want to design a

[06:45:00] study with a family-wise Type I error rate of 5%. For moderate correlation between over-survival milestone survival between .5 and .8. That is the shaded area over there. Let's say that I split this by four. I want to allocate three-fourth of alpha to over-survival, and one-fourth to milestone survival. What happens is that, when you have a correlation between .5 and .8, you still have at least .04 Type I error rate

[06:45:30] allocated to over-survival. In this case, as an example, if I have the correlation of .7, I am allocating .042 to over-survival, and .014 to milestone survival. As you can see, they add up to .056. The additional .006 comes from the correlation of .7. Now I have the milestone survival alpha of .014. I am going to use this to design my study.

[06:46:00] Sample size determination. This is the last step. End point use and sample size, as

we mentioned, milestone survival at two years, in this case. Treatment effect improvement in two years of 20%, going from 31% to 51%. Type I error rate of .014, power 90%, assuming that I am going to accrue 34 patients per month. What this tells us is that I will need a sample size of 342 patients. We are talking about patients, not number of events. We need 342 randomized patients. Accrual
[06:46:30] duration is going to be 10 months, follow-up duration 24 months, because that's the ideal follow-up duration. That's where the 31% and 51% came from. That duration is determined by milestone survival and it mitigates the issue of censoring, if you ensure that the follow-up actually reaches 24 months.

What is the timing of the analysis? As I mentioned earlier, it is going to be two years after the last patient came on to the study, or approximately three years
[06:47:00] after the time of the first patient on the study. If we truly use this approach, I would think that two years is actually a much better criterion, because you never know that it is going to take only ten months to enroll patients. It could be faster; it could be lower. You want to make sure that patients do indeed being followed for two years.

When we analyze the over-survival end point, the power can range from 95% to
[06:47:30] 90%, from no delay to eight months delay. Again, this is just one set of examples. The survival group doesn't have to follow exponential distribution. You can make them look whatever shape that you want. This is just an illustrated example. I use the curves on the right to show you that in this case, in particular, if I design a study with 340+ patients, by the time that I reach two years of follow-up, the power is going to be 95 to 90. if I analyze data on over-survival end point using a Type I error
[06:48:00] rate of 4.2%. I analyze the milestone survival with the rest of the 1.4%.

What are the advantages of the proposed design? It is simple. Number two, over-survival remains an important end point. There is no assumption in the shape of the over-survival curve. I don't have to assume what is the delayed duration. I don't have to assume what is the cure rate. Number three, as I mentioned, the design is agnostic to uncertainty of the delayed duration and long-term survival. I can predict
[06:48:30] the timing of the analysis. I incorporate the assessment of the tail of over-survival curve, because I did allocate 1.4% to test the milestone survival at two years. Milestone survival analysis can capture a unique lay separation with a higher bar, assuming, hypothetically, in your analysis your over-survival is not positive. Miss 4.2%, but your milestone is positive at 1.4%. It is telling us something. If in your
[06:49:00] entire curve, entire analysis, over-survival is in a positive, there must be a cross or whatever at the beginning, but for some reason you have a tail, that is so good. That is significant at 1.4%. That is something that is extremely important, in my opinion.

Minimal loss of false positive rate in over-survival testing. It mitigates the risk of inflating a false negative rate. In summary, it's a simpler design for future late state IO studies using milestone, by allocating one-fourth of false positive rate, under
[06:49:30] moderate correlation, the probability of success can be assessed under plausible over-survival assumptions. We also incorporated a clinical meaning for milestone

aid improvement into the study design. If it is so desired that we want to look at the data earlier, for example, although I'm interested in two-year milestone survival, but I want to analyze the data 18 months after the last patient is randomized. What happens is that you are going to have a group of patients with
[06:50:00] administrative censoring. In this case, you just need to inflate your sample size. There is a way to do that.

Section 41 of 45 [06:40:00 - 06:50:04]

Section 42 of 45 [06:50:00 - 07:00:04] *(NOTE: speaker names may be different in each section)*

Speaker 1: In this case you just need to conflate your sample size. There is a way to do that, because you know you need to take into account the variation in the study design. So, instead of 340 plus patients, you may need to heave 60 or 70 patients. Finally, the proposed studies [inaudible 06:50:15] clinical meaningful information and statistical simplicity. I believe that is my last slide and a list references. Thank you.

[06:50:30]

David Rimm: I'm impressed. It's 4 o'clock on a Friday afternoon and you all stayed to listen to a pathologist. That's pretty good. SO hopefully I can make this interesting for you and
[06:51:00] even maybe potentially make it reasonable quick. First, my disclosures here, you can see that anything that I do in my lab can't effect patient care unless I have partners, and these are my partners.

I want to talk to you about 4 specific things today, about critical components for companion diagnostic assays in immunotherapy. Some of these things also apply outside of immunotherapy, but I think these are the 4 things that potentially can help you as you design your biomarker piece of the clinical trials that you're doing.

[06:51:30] The first is to control the pre-analytic conditions. For this one you kind of get an 'A', there has been a lot of awareness as this study is back from 2010, and what they showed in this study is that if your estrogen receptor rate in breast cancer was more likely to be negative on Friday then it was on Monday. Why was that? That's because it sat over the weekend. Because it sat over the weekend the cases that
[06:52:00] were positive decayed and became negative, so you have a higher negative rate on Monday, or an increased probably false negative rate as well on Monday. This was actually observed and over the period of the last 5 or 10 years people have become aware of this. Be sure that when you take your specimen, you take it in a way that it can be rapidly prepared which is usually a core biopsy. Core biopsy specimens, or rapidly fixed specimens as opposed to resection specimens should be used in your trial design.

[06:52:30] Also, stain the slides immediately after they've been cut, there's a fair bit of data suggesting slides oxidize or actually hydrate, and the hydration effects antigenicity. That's another thing that's recommended.

Finally, choose targets that are stable to warm and cold ischemia if you can't do

that, because that's what happens when you resect the [inaudible 06:52:51] and the tissue sits around in the OR refrigerator, you lose antigenicity.

[06:53:00] Next thing, not so carefully done but quantitatively tighter are monoclonal antibody for IHC assays. Historically, IHC assays were even done with polyclonals. Now I think most antibodies that are out there are being done with monoclonals which gives you reproducibility but they aren't all done at the optimal signal to noise. This is the most recent and graphic example of this.

This is from the blueprint study led by Fred Hirsch, who many of you are familiar with, looking at PDL1 in lung cancer and you can see that these serial sections that
[06:53:30] this particular assay, not the antibody, but the assay, recognized about half as much as these others. They claim that it was because they were trying to emphasize the immune cell component but this is biochemistry, it's not cell biology. The laws that govern biochemistry have to do with affinities of antibodies and don't care which cells they're in. There's no way to actually emphasizes one over the other unless it's a secondary thing. The antibodies themselves will bind to the antigen independent of the cell type in which they're located.

[06:54:00] This is from the NCCNBMS version of a comparison study and so in a second study that was statistically powered 13 pathologists at 7 different institutions looking at 90 cases found the same thing which was found in the blueprint, because the antibodies were probably not titered properly. Here's the summary data of that, this was presented at [inaudible 06:54:22] Chicago just about a month ago by
[06:54:30] myself. You can see that this is the 142 assay versus the other assays here that you can see are done differently. Even though we're trying to detect the same molecule or theoretically detecting the same molecule. This was true whether it was in the tumor cells or in the immune cells. That actually makes a lot of sense since the antibody antigen interaction is independent of the cell type.

Here's a summary of that data, you can see that depending on how you choose the cut points, the 50% cut points are in blue, the 1% cut points are in red. You
[06:55:00] dramatically decrease the number of patients that might go on to a trial or might qualify for a trial, if the antibody isn't optimally titered. You can see the distributions here above.

How do you optimally titer an antibody? Well you have to have a whole series of ... The best way to do this is actually to calculate the signal to noise. This can be done with a series of specimens as in an index array here, that's done quantitatively. You can see when SP142 was managed in a quantitative fashion, the high cases, the low
[06:55:30] cases, and then the signal to noise shows you a peak and an actual concentration at which you can optimally measure PDL1.

Since I have a relatively limited time and you're not an audience of pathologists, I'm going to move on to the next point, which is how to do the analysis method. This is another point that I would encourage you ... It's not 1995 anymore, we now have lots of machines that we didn't have back in 1995 to measure things more carefully.

[06:56:00] Choose an analysis that can either be achieved by a pathologist. Pathologists are actually pretty good at what we do, but there are some things that are better done by machines than pathologists.

Here's an example of that, this is a study that we recently published comparing a pathologist to a pathologist, a semi-automatic method to a semi-automatic method, the quantitative fluorescent to quantitative fluorescent, and you can pretty clearly see the difference. Pathologists tend to group things, we tend to
[06:56:30] group things as high or low, even though this is a continuous population, and the DAB, diaminobenzanine, is a chromogen. Chromogens are harder to measure because they absorb light as opposed to mechanisms that emit light.

Can a pathologist be used to score? Absolutely, in fact we saw that. We looked at pathologists agreeing to ... Scoring all these different antibodies from the NCCMBMS study, the reproducibility, the interclass correlation coefficient for
[06:57:00] pathologists for each of the antibodies was above .8 or close to above .85. That's really good, that means we agree really well. However, for immune cells we were down in the .1 - .2 range. That means we can't do that, so if you need to measure something in immune cells you should measure it with a tool other than pathologists. If you need to measure the PDL1, in this case, in the tumor cells we were very competent at doing that.

Now what about the cutoff? The cutoff makes a difference as well. You can see our
[06:57:30] concordance .75 to .77 when we looked at greater than 50%. But dropped to .53 or .61 when we used a cut off greater than .1%. This suggests that not only do we need to think about the assay, [inaudible 06:57:39] task of the pathologist, but we need to think about what we're asking them to do. If we're asking them to do something that's pretty straight forward, like greater than 50%, or less than 50% we can do a good job of that. But if we're asking them to do something hard like greater or less than 1%, we have a little less concordance in that assay.

[06:58:00] How many pennies are here? 5 right? No, it's actually 100. This is the reason that you shouldn't use a chromogen when you measure things, because chromogens absorb light, and so you can't really see. This is the equivalent of a chromogenic assay, this is the equivalent of a light emitting assay, like florescence. When you think about which assay to choose , I would propose that you choose a fluorescence based assay in order to be more continuously quantitative. The light
[06:58:30] bulbs don't illustrate the point as well, but the power of three light bulbs emit 3 times as much light not one.

These are examples of assays that are under development now, that I think we might see more of, which are combining molecules using multiplex fluorescence. There are now 140 Perk and Elmer Multiplex vector machines out there. So the argument that nobody can do it but you is no longer valid. There's lots of machines
[06:59:00] out there that can do quantitative fluorescence. This assay we'll show at World Lung, is actually very provocative with respect to power to predict response to PDL1 therapies. Others may be similar, and we've also seen data from Janis Taube

and from the Toni Ribas lab suggesting that the quantitation of the immune cells might be as valuable or more valuable to predict response beyond just looking at the tumors themselves.

[06:59:30] What about other things you can do? RNA signatures are a very provocative concept and they give you the kind of reproducibility, because you're using a machine as opposed to a pathologist. Now I love pathologists , I am one, but there are times when it's appropriate to use a machine and I think some of the preliminary data we've seen and some of the things we've heard about from Merk and their ATG panel has done on the nanostring platform could be another
[07:00:00] direction to think about as you design your assays. This machine is moving into Cleo labs, there is Cleo lab approval for some of the tests in this box ...

Section 42 of 45 [06:50:00 - 07:00:04]

Section 43 of 45 [07:00:00 - 07:10:04] *(NOTE: speaker names may be different in each section)*

David Rimm: CLIA Labs. There is CLIA Lab approval for some of the tests in this [Bat 07:00:03] Box, and I predict that we might see a lot more as we need to have higher level more complex types of analyses than [inaudible 07:00:11] platform. I'm not sponsored by them, I don't have any funding from them, but I do think that it's a very interesting platform for potential measurement of RNA as a companion diagnostic test.

It wouldn't be complete to not talk about mutational loads. I want to just briefly
[07:00:30] mention that. The [Risby 07:00:29] Group and others show that mutational load actually looked at as a promising mechanism to tell it apart, and DNA is the easiest thing to look at. You just sequence it and it's really stable. You can't ... It takes weeks or months for it to go bad, or years if you want to look at dinosaur DNA. It looked like there was actually pretty good promise, but it's actually not. We've actually replicated this and this is some data from Yale, and we got pretty much the same results-that is statistically significantly increased numbers. Mutations are
[07:01:00] associated with response as opposed to non-response and the same thing with Class I neo-antigens. Both of these tests have a sensitivity of only 68% and a specificity of 58%. In bio-markers, it's all about sensitivity, but you also can't forget about specificity. I would argue that this test has neither the sensitivity nor the specificity to become a companion diagnostic test.

[07:01:30] Some companion diagnostic tests like HER2 in breast cancer has extremely high sensitivity in the 90%-95% range, but their specificity isn't very good but they got away with it. However, we just saw an illustration in the field of immune therapy where a sensitive test failed because it wasn't specific enough. That was the 5% cut-point for BMS, whereas the specific test greater than 50% cut-point showed a
[07:02:00] response that would probably change care. Don't forget about sensitivity and specificity.

Finally, the last thing that at this meeting may be appropriate to talk about, it's really hard to standardize resist criteria. It's almost impossible and that's what you all have spent the last two days talking about. For an assay it's not so hard. We're

actually doing something that's the equivalent of measuring glucose or sodium in the blood. We should actually be able to rigorously standardize that, and we're working toward this right now. What we would suggest is using standardized
[07:02:30] isogenic cell lines with known amounts of PDL-1 expression so that this standardization array can be run every day. We are now in the process of figuring out the exact concentration of PDL-1 in 15 different isogenic cell lines in collaborative project with this company called Horizon Discovery because they seem to have most of the IP locked up around isogenic cell lines.

This would allow us to test every day to run that stainer, and every time we run it, know that our stainer is correct and that we're actually quantitating. If you're
[07:03:00] above a specific amount, then you can call it positive. It's critical to determine what we call positive and what we call negative and we can't have that shift, and so use some sort of standard. It is possible to standardize even tissue on slides, even measurements of tissues on slides. This is how I would recommend doing that.

With that I'm running out of ... I've left 30 seconds for questions, I guess. Thank you very much.

[07:03:30]

Sridhara: I request all the speakers to come to the panel table. First, I would like to really thank all the speakers. I know it's always hard to give presentations in the last
[07:04:00] session. They were very interesting and very informative. Of course, this leads us to what is the way forward, and I want the panel to weigh in on some of these in designing future trials.

For a moment, let us think about in the progression, definition of progression we have considered resist everything, but from what I was hearing today and
[07:04:30] yesterday, including this pseudo-progression versus other true non-pseudo-progressions, can you clinically differentiate between these two and can you standardize the clinical judgment into having it as a component of defining progression itself. To my clinical colleagues and the panel.

[07:05:00]

Female 2: [inaudible 07:05:34] would be me.

I don't know the answer to that. I think we struggle, you know, looking at a CAT Scan it's ... I mean we're looking at some [radio-mic 07:05:11] features to see if we can see after a large data set whether we can use not just size as our predictor of response, whether you can look at texture and other features to predict that. I don't know if there is a way to really know who is a true progressor and who is a
[07:05:30] true pseudo-progressor. I'm sure others in the audience have thoughts.

Male 2: It is also defined a little clinically if the patient looks well, feels well, and has a bigger lump, it's much less likely progression. If the patient looks terrible and the lump is bigger, it probably means something.

Female 1: [07:06:00] I think where we are struggling within the regulatory agencies, we need some kind of objective measure so that there is no bias that's introduced, and I think we have to differentiate between what is medical practice and when we are talking about clinical trials. In the medical practice, yes, you know you can wait or do differently, but here we have to really look at it objectively and we need to bring some kind of standardization. Yes?

Male 3: [07:06:30] Dan [Chin 07:06:21] and I were talking about this over the break and I think that there is an opportunity certainly in neoadjuvant setting and maybe in window-type settings to define molecularly progression or not progression. I don't think we've taken ... I think we're beginning to take advantage of that opportunity where we look at post-neoadjuvant or post-window therapy specimens and say, "Okay, now let's follow that patient along, now let's go back and look at the biology, and is there a biological correlate whether it's a multiplex signature that is associated with progression or not." I think that there's an opportunity out there to become more [07:07:00] quantitative that define progression in ways other than radio logic.

Female 1: [07:07:30] I think we entered a little bit into adaptive trial designs, looking at some platform trials or even otherwise, you know. We have talked about this non-proportionately or at the beginning that there is no difference in the survival curves. Would an interim analysis be good to not stop the study, but make adjustments in the analysis, as well as the sample size, to see if this should be different? This is for my statistical colleagues here.

Male 2: [07:08:00] [Rami 07:08:45] in my opinion, we have to do this carefully because if you really conduct your interim analysis when the curves are still together, let's say with a little bit if separation, small delta, then you do your symbol size re-estimation, then you increase your symbol size, then what happens is that you are going to have more patients with a shorter follow-up.

[07:08:30] Number 2, you may need a huge sample size that you may never finish your study because by increasing the sample size, let's say that originally you planned to have 200 patients now all of a sudden you say that I need 500 patients. It takes even longer and with longer studies and patients with shorter follow-up, I'm not so sure if this is going to help, although I can see the benefit, but we have to do this carefully.

Male 1: Raj, were you talking about increasing the sample size in terms of sample size re-estimation or were you talking about changing the analysis plan?

Female 1: Changing the analysis plan ...

[crosstalk 07:08:54]

Male 1: Based on an interim analysis.

[07:09:00]

Female 1: Let's say that whether it is ... Because we don't want to look at all of the data and then say it doesn't fit non-proportionality and use some kind of other models that we have discussed right from this morning, or can we do it prospectively somewhere in between and say we are observing this phenomenon.

Male 1: I think that is a great research plan for [Ty 07:09:22] to work on the next few years.

[07:09:30]
Female 1: Another question is, would that interim analysis give us some information, provided we have some follow-up on at least some reasonable number of patients whether a particular sub-group based on some of the molecular bio-markers, or if we are able to get better understanding of PDL-1 using the [inaudible 07:09:53]
[07:10:00] companion test that was discussed by Dr. [M 07:09:57]. Whether we can use some of that and see who are these patients, and ...

Section 43 of 45 [07:00:00 - 07:10:04]
Section 44 of 45 [07:10:00 - 07:20:04] *(NOTE: speaker names may be different in each section)*

Speaker 1: ... that and see who are these patients at the end of it, and then modify the design.

Speaker 2: There's been a lot of statistical work done on that, but you have to pre-specify at the beginning what you plan to do, what you're going to look at, and what your adaptions are going to be in terms of the analysis. You pay a price for it, of course, a little bit, but you can do it.

Speaker 1: Now the question is for all of you. You have heard about different summary
[07:10:30] statistics. I don't think anybody came to the conclusion what is the best test, but everybody is looking still at log-rank test. Let's say we agree that there is a treatment effect, and then we are trying to summarize. As a clinician what would make more sense for you to report, and as statisticians, what would be the best
[07:11:00] way if there is non-proportionality? I'm just thinking about future designs, how can we do this, and how should we plan, so it's in that sense that I'm looking at it. No takers?

Speaker 3: I actually like Dr. Chen's design, and I wanted to ask him whether he can envision using something like that for a randomized Phase II study, or is that too hazardous to apply to? Does it have to be a late-phase study?

[07:11:30]
Dr. Dan Chen: I think the reason why we have Phase II studies is because we're going to get some signal earlier. I think that if you look at this, it takes time. The example that I gave you, it takes 3 years, and that's 300th...

Speaker 3: Can you take an earlier milestone, like a year?

Dr. Dan Chen: Yes, and that's why at the beginning of my presentation I said that I was not going to resolve these issues because if 1 year is a meaningful time point from a treating
[07:12:00] physician's perspective, that is clinically meaningful, and I think that is 1 year, but

sometimes 1 year is just too early. If you happen to pick a time point that is still ... there is a delayed treatment effect, and I think earlier we talked about in the future the control is going to be immunotherapy. So all of the arms contain immunotherapies, and what happen is that you're going to have the delayed treatment effect that potentially is getting longer and longer, and I don't know ... In one of the opening slides that I had that I borrowed from Dr. [Urba 07:12:26], if you actually have a curve that is all the way up there off the yellow curve, I don't even

[07:12:30] know what type of separation that you can afford to even to show treatment effect because you don't have any room any more.

Speaker 1: But again, with the milestone, whatever time you fix, then you have to have it that everybody is followed up to that point. You cannot have any censoring before that, otherwise it becomes difficult.

Dr. Dan Chen: Right, and I think probably ... and also briefly mention about that as well is that

[07:13:00] unless that you are willing to ... I'm just saying that your target is 2 years, and you want to analyze data 6 months after your last patient's even [inaudible 07:13:07]. I'm not talking about that. I'm thinking about if your milestone is 24 months and you're thinking about analyzing the data 18 months after, meaning that you still have ... you already have a large proportion of patients being followed for 2 years, and the remaining patients who are censor will not change the clinical interpretation, then obviously you have to take a hit, and the hit will means that

[07:13:30] you have increased the sample size.

Speaker 1: Yeah. Please.

Speaker 5: Thank you, Dan Chen, Genentech and Roche. I wanted to make a comment and also ask a question related to Dr. [Rim 07:13:42]'s very interesting and insightful findings related to PD-L1 IHC antibodies. The first place I wanted to start is an area where I think there's a lot of confusion in this field, and that is in the last year

[07:14:00] people have started to talk about IHC antibody performance in the same vein as assay performance. Whether or not, I don't think anybody today understands whether the different antibodies, their performance is due to either greater sensitivity or greater specificity on one versus the other antibody, but when it comes to assay performance, I think that many of the conclusions out of antibody performance are now being transposed onto assay performance.

[07:14:30] An example of this, when we look at SP142 for immune cell scoring with trained pathologists, we have concordance rates of greater than 90%. We've shown this repeatedly, so you can reliably score with an appropriate assay immune cell PD-L1 staining. It just requires a certain amount of training and an assay that is tuned to be able to do that. I think that's a really important distinction between antibody

[07:15:00] performance and assay performance. Similarly, when we look at scoring by the different IHC assays, they score very similar or you could approximate the same patients, for example, at the 00 level and possibly at the very high level as well.

Again, coming to antibody performance, where there may be some differences due

[07:15:30] to specificity or sensitivity, we don't know, but in terms of assays, these assays, as they're currently appropriately designed and scored, appear to delineate the same groups of important patients. We've taken a further step to try to understand this, so as you yourself mentioned, because IHC can be subjective, it's so important for us to utilize something that ultimately is as close to a gold standard or something that's comparable in this field. I think you mentioned gene expression and T-cell
[07:16:00] effector signatures as examples of that.

When we saw these differences in antibody performance, as noted in Blueprint and your own publications, we immediately went back to look at gene expression. Gene expression correlates very well, whether it's PD-L1 gene expression or T-effector
[07:16:30] gene expression with SP142 staining, or SP142 scoring as an assay. I guess when we talk about this field and taking it forward a step, if ultimately T-effector signatures or PD-L1 gene expression ultimately score the same groups of patients as we generally see with PD-L1, how will we look at this as a field? Will we still be of the belief that PD-L1 IHC was a sub-standard assay and we just needed to develop
[07:17:00] something better, or will we come to the conclusion that immunity is complicated and ultimately we're able to sub-segment patients that are more likely to respond and less likely to respond, but we're never really going to be in the position we are, for example, with a V600E mutation and a BRAF-targeted therapy.

Speaker 6: Can I just respond to that? I think Dan and I are very much on the same page with
[07:17:30] respect to having quantitative and more complex assays, expression profiling, et cetera, and I completely agree. But, I disagree completely ... the data that I showed you was all assays. Those were all assays performed as close as possible at the time they were performed by the instructions of the manufacturer Ventana or Dako. That work will come out in publications shortly, but those are the different ... The antibodies actually perform identically, and that came out in work from my lab in
[07:18:00] JAMA Oncology on August 18th. The antibodies are the same. The antibodies are the eggs, the assays the cake. How you make the cake, the recipe that you use, dramatically affects what the pathologists score.

What I showed you was the results of 13 pathologists all scoring at their own ... scoring the same cases, the identical cases. That inability of the pathologists to have concordance on measuring of immune cells is ... That's what we found. Now, I know that you can make an argument that, "Well, they were untrained." That's
[07:18:30] absolutely true, as are most of ... You know, we were all trained. We're all board-certified pathologists. Did we get specific training? No. Specific training has uniformly failed. There's not a single assay that we do that requires specific training that exists in the armamentarium pathologists use today. We rely, the 18,000 of us across the country, rely on the training that we obtained during our residency and
[07:19:00] subsequent CME to ultimately judge these kinds of assays.

While it may be desirable to have some special training for some special assay, it's unlikely that that will be a successful pathway. I would argue that the data that you saw represents the concordance that's statistically powered, prospectively designed. It's Level 1 evidence for comparison of assays, not antibodies but assays.

We're reporting the data. There's not ... It is what it is.

[07:19:30]
Speaker 6: We agree that those are important findings, but I guess where we would come in is I think training ultimately is important and especially if you're going to look at different assays that are going to appear differently. Those analytic cutoffs ultimately are critical to the performance of any given assay, especially one that's as subjective as IHC. We recognize that as part of the platform issue for IHC. That's
[07:20:00] one of the reasons why we've pushed so hard into trying to look at gene expression both for PD-L1 and for effector signature-

Section 44 of 45 [07:10:00 - 07:20:04]
Section 45 of 45 [07:20:00 - 07:34:06] *(NOTE: speaker names may be different in each section)*

Speaker 1: Gene expression both for PDL1 and for factor signature because at least you are able to remove some level of reader variability.

Sridhara: Can I go to three? You have a question? Yes.

Janaki: Hi Janaki Veeraragavan FDA. As a biomarker person again, I want to address this to Dr. Rimm. You recommended use of core needle biopsies as the most reliable
[07:20:30] specimen type for development of biomarkers. I want you to address the issue of false negativity due to poor tumor or immune cell representation in these specimens. How would the field address that issue?

David Rimm: That's the issue of heterogeneity. That's a huge issue that we are still in the process of understanding. How much of a sample do you need? The current standard of
[07:21:00] care is less than 0.01% of the tumor. That's how much we sample. We sort of think that's enough because historically it was enough when we were just looking at morphology. The question is what is now the current, what should be the standard of care for representation of tumor and filtrating lymphocytes for example.

Nobody knows the answer to that question. We've looked at it a little bit by looking at different cores in larger breast cancers and that was published just recently, where we find that the heterogeneity foretells at least in breast cancer is on the
[07:21:30] centimeter level or the millimeter level that is within the fields of view. If you look at serial sections, serial sections are only five microns apart. They have a quantitative identity of 0.99 or 0.95. If you look at three different blocks, that is different regions of the tumor that's the centimeter scale.

They actually agree about 66% of the time, but if you look at random fields of view, that's where the variation is. I don’t know how we actually figure out. We were
[07:22:00] wrestling with what is the minimum amount you need in order to represent that randomness. I don't think we have an answer to that question yet.

Sridhara: Yeah next question.

Shahul: This is [Shahul Fenn 07:22:09] medical reviewer from CDER FDA. Thank you for an

excellent panel discussion presentation. I have a question actually for four of you except Dr. Siu other than a comment, which is a fantastic presentation. If I started
[07:22:30] with a first question to Dr. Korn. I know that Dr. Rimm in detail talked about the assay design for the testing assay. My question for Dr. Korn is as a general investigator, to develop the biomarker assay where you think that sensitivity or the specificity is good enough to go in because, when you go into the trial to separate
[07:23:00] the positive or negative, then what about the reproducibility or validity or do you engage the CDRH or where you cannot do that, that's what I'm wondering.

Speaker 2: Maybe too much.

Ed Korn: Yeah I hope we are going to do it before phase III. You are going to do it in phase II, you are going to do it in phase I. You are going to do it while you are doing phase I and phase II. Once you get to phase III, I hope you have something that's analytically valid, that's reproducible, and it is measuring something. Now what the
[07:23:30] cut point should be, if you ask me about a cut point for an assay that you know is kind of analytically valid, which sounds like just some questions about that, that's difficult also.

But again we would try to do it in phase II as much as you could because once you get in phase III, you are into a large sample size and it is going to be hard to design it, to look at the lot of different cut points at that point.

[07:24:00]
Female: I guess the dilemma is, at the beginning you take quite a while to really validate that assay. I was wondering what's the procedure, what's some kind of tips you could give for the general investigators for that part.

Ed Korn: I don't have any advice to it. I think people are still arguing about how to measure it too. It can take a long time.

Shahul: Thank you. I have a question for Dr. Kirkwood. I really enjoyed your presentation
[07:24:30] about the combination, but I'm a little bit disappointed about it. You didn't mention about another biological product is [Imulitic 07:24:36]. Actually that was my baby product at CDER. My question is because actually into tumor injection specifically, if you look at that Imulitic approval, and there is a subgroup actually is in this part. They have survival difference.

[07:25:00] I would like to see your insights or your thoughts on whether or how or why not that much interest in tumors of combination with Imulitic and others?

John Kirkwood: In the 12 months I had, I couldn’t hit all of the things we are doing. We are part of the Masterkey T-VEC trial, so we are as interested in the impact of intra-tumoral T-VEC injections. I think TLR9 agonists, the checkmate agent equally interesting and
[07:25:30] also under study, many I think such approaches make good sense to try to stir up the tumor, to take the cold tumor to hot and to make greater impact with the CDI agents, the IO agents we've been talking about otherwise.

Shahul: Can I ask, I think there is another question? No. Okay.

Sridhara: We are running out of time. We have one minute. Okay she has a question.

[07:26:00]
Female: I just wanted to make a comment. This is going back to the assays, and it is really unfortunate the CDRH isn't here. It would've been wonderful to have your colleagues from there at this last discussion. Just fascinating. I think we talked a lot about validating an assay and we talked about cut points and reproducibility and how specific it is. That's all wonderful, but we have to remember validation is not just analytic validation.

It is also clinical validation, and I do want to go back to a very specific term. It is a
[07:26:30] companion diagnostic. It is not just a diagnostic that's in vacuum. It is a companion to a drug. These are relative to the performance of the drug and the whole point of this exercise is to identify the appropriate patient to get treatment by the appropriate drug. I agree with you that there is a lot of confusion in the field out there. We are still grappling with is it the right biomarker, is it the right level of expression?

[07:27:00] Now we are going from one protein to signatures and so on and so forth. From a clinician's perspective, at the bedside from a patient's perspective, they are not sitting down and going Did I get SP-142? Did I get 22C? They are just saying do I get ... They don't even care about the test. They just want to make sure they get the drug. It is our responsibility as a community to figure out how we resolve these issues so the patients get the right treatment.

[07:27:30] I do think community efforts and efforts like the blueprints and full disclosure, I was very much a co-conspirator, but that was really done keeping in mind all of the realistic challenges that were ongoing at the time for very, very competitive development programs. For direct competitors, coming together to say what can
[07:28:00] we do to make this easier for patients? That was my piece.

Male: I can just respond to that. I think that I agree with you 100% that in fact we want to keep in mind that these are assays that have to have a direct connection to the drug if they are in fact to be a companion as opposed to a complementary diagnostic assay. That companion suggests biology, so it is really important that when we name an assay, we are measuring the biology, and so if that's the case
[07:28:30] then all the assays should measure the biology the same.

If the assay is different, that means that one of the assays or more is failing to accurately reflect the biology. One of the problems I think with PDL1 when these assays were approved is they were approved without respect to accuracy, because there was not a mechanism to accurately determine exactly how many molecules of PDL1 represent on each slide. Instead, we used a surrogate for accuracy which was pathologist's opinion.

[07:29:00] Unfortunately that pathologist opinion didn't result in the four assays being equivalent, which if you did mass spec on all those tissues for example or something that was a little more rigorous, the assays would be reflecting the biology as opposed to reflecting the result of the assay.

Sridhara: Thank you all very much. I think it was a very good session. I'm sorry we couldn't get to all of the questions. Of course you can always come and approach the speakers later.

[07:29:30]
Female: I think I can't let that last statement go unchallenged in fact. All of the tests that were approved as companion diagnostics have been evaluated for their analytic performance characteristics including accuracy. I don't think that ... I'm talking about the PD1 drugs. I'm not talking about, but the companion diagnostics for the PD1 drugs were evaluated.

Male: How?

Female: I'm sure we can find it on, published on the website?

[07:30:00]
Female: Yeah.

Female: We will be happy to send those documents to you.

Male: I've seen them. Thank you.

Sridhara: With that, let's end with ... I think final remarks will be given by Dr. Hazarika.

[07:30:30]
Dr. Hazarika: Good evening everyone. As they say, time flies when you are having fun. I think we can all agree that we had a fabulous two-day immuno-oncology drug development workshop. We had a whole lot of discussions about ongoing challenges. We heard
[07:31:00] the criteria needs to be changed, standardized criteria. We heard about RECIST, I RECIST, IM RECIST. We understand that the RECIST criteria may not be optimal and therefore, we do need to find some standardized criteria.

We also heard about endpoint challenges and preferences. Durable response rate or overall survival, but with different statistical analysis. We heard about the
[07:31:30] research from my FDA colleagues about the association, the weak association between PFS and OS, but the potential moderate association between an intermediate endpoint and OS, so many of these concepts will just continue in future research. To quote Dr. Sridhara this morning, we don't have all the answers.

[07:32:00] We hope to continue the discussions, which we started today and to move the field forward with alternate endpoints, alternate criteria, and the novel study designs,

[07:32:30] most of which are already ongoing. I would just like to thank all the speakers, the panelists, the moderators, for their fabulous presentations, and excellent discussions. Thank you all for coming all the way from all the ends of the east, west, south, north coast and helping us make this workshop a success.

[07:33:00] I'd also like to thank the audience and all the currently 1370 enrollees that we have out here as well as online and for all your interest and your questions and interactions, which just made for a very lively discussion. I would like to thank my coaches, [Dr. Theoret, Dr. Topalian, and Dr. Wolchok 00:13:04]. It was an absolute pleasure and maybe we can have another workshop soon. Finally, I would absolutely like to thank Anna and Nicole from ACR, who made all this go very smoothly and Beverly from FDA who coordinated everything. Thanks very much and have a lovely weekend.

Section 45 of 45 [07:20:00 - 07:34:06]