6-13-16 FDA-AACR Oncology Dose Finding Workshop – Session 3 Transcript

Eric Rubin: If everybody could please come back in and take your seats. It's been such a great meeting and we've had such good discussion. We're running well over and, like myself, I think several folks in this session have hard stop at 5, so we're going to have to try to find a way to make up some time. This session is on designs for dose optimization studies, premarket and postmarket. We've talked a bit about that before. You're going to hear some great talks that are going to go into some more detail and then we have a panel discussion. Again, for the essence of time, I won't introduce each speaker. Please introduce yourself and keep in mind you have 15 minutes. I've noticed some of the decks ... there are more than 15 slides. Feel free to skip a slide, if you want, so that we can get to the panel discussion. Thanks.

Haleh Saber: Good afternoon. I'm Haleh Saber, Deputy Director in DHOT, Division of Hematology, Oncology, Toxicology in OHOP. This talk is on immuno-oncology products and, mainly, on first in human dose selection. Because there is so much interest in this area, we decided to look at INDs and see what we've learned so far and answer certain questions around first in human dose selection. Are we too low? Are we too high? Is there a safe approach for all the INEs we looked at and is there a need to optimize phase 1 clinical trial design?

What is MABEL? It stands for minimally anticipated biological effect levels. If you hear a MABEL dose, that's a very low-dose with some anticipated effects. If you hear a MABEL approach, this is not really a single approach. It could be selecting a dose based on any of the following: xenograft data, activity data, binding data, and other data as applicable. It's used for immuno-oncology products. It's sometimes also used for other products when there is no pharmacologically relevant species to do a tox study.

What led to MABEL? That was in 2006 when the antibody against CD28, TGN1412 was given to healthy subjects and resulted in cytokine storm and organ failure. The dose was a 500-fold below the doses that were given safely to animals. It was later found out that that dose actually resulted in 90% receptor occupancy. That basically prompted first in human dose for immuno-oncology products at 10% receptor occupancy.

Since then, we moved up a little bit. Now, the preference by many reviewers in DHOT and also by many sponsors is 20% receptor occupancy or 20% pharmacologic activity. There are various methods to come up to estimate the first in human dose based on pharmacologic activity or receptor occupancy, but for the purpose of our project, we used two simple equations and these are based on Hill equations and [inaudible 00:03:30] kinetics. Basically, if you want to keep your receptor occupancy at 20%, you have the dissociation constant, then the only unknown is Cmax, the plasma concentration of your drug. If you want your pharmacologic activity to be at 20%, then you have [inaudible 00:03:48] activity EC50 and the unknown is, again, the plasma concentration, and you convert that to a dose.

We collected the data. I'm going to keep this slide very brief. Collected data on immune oncology products in DHOT and we wanted to make sure that these products have the potential to activate the immune system, so our products included antibodies that were checkpoint inhibitors or checkpoint stimulators, and also CD3 by specifics. We also included some other products with the target potentially being involved in immune system activation. We excluded interleukins and fusion proteins. Also, we want to make sure we have sufficient clinical data. We had 31 INDs, 5 of them by specific constructs. These are just the list of some of the targets of the antibodies.

We collected pharmacology data, more specifically in vitro binding and in-vitro activity data to use it in the equations I showed you on slide 6, toxicology data, and we also collected phase 1 clinical trial design of what was the first in human dose, what was the rationale for the dose, was it a 3 plus 3 design, was it a single patient cohort, was intrapatient ... those escalations proposed are allowed. We also collected some data on treatment for CRS, cytokine release syndrome and infusion related reactions, not so much as what the treatments are, but rather was it done prophylactically or was is done after the occurrence of the first event.

Independent of what the sponsor proposed for first in human dose selection, we came up with first in human doses that matched with 20 to 80% receptor occupancy and 20 to 80% pharmacologic activity. Why did we choose those 2 equations in these 2 approaches only? Because we had the KD4 almost all INDs except for 2 of them. We had the EC50s for most INDs, so basically, we could do a cross IND comparison. We did have the data for TGN1412, so we were able to actually also compare it to that product.

There are difficulties associated with other methods. For instance, for xenograft ... you heard a lot today about xenograft studies. Not all sponsors did the studies and whenever it was done, it was not homogeneous. Sometimes it was with a clinical candidate, sometimes it was with a surrogate, so again, we wanted to stay away from those and make sure we have a homogeneous way of addressing this.

Preliminary conclusions. Some initial conclusions, I'm going to show you. Right now, at this moment, we don't have any conclusions for by specifics because we didn't have enough of them and they are not homogeneous. Some of them are given weekly. Some of them are given daily. Some of them look more like fusion proteins. They don't always look like antibody in structure.

What about antibodies? For half of the INDs that we looked at the dose, the first in human doses are very low, extremely low, like 100,000 or even 10,000-fold below the doses that have been given safely to patients. What approaches used by the sponsor for first in human dose selection? A variety of approaches based on xenograft data, based on PK modeling, based on binding data, activity data, and so on, and some sponsors actually didn't have any explanation. They just presented the dose, just said that this is a MABEL dose.

These are all the MABEL approaches used by the sponsors, but some other sponsors also compared their products to approved products or to products in development, and they were actually to start at a dose that was higher than what is really a MABEL approach.

When we independently look at the first in human doses, we concluded that 20% receptor occupancy. A first in human dose that results in 20% receptor occupancy is actually very low for all the antibodies, and again, we are not talking about by specifics, just antibodies; 20% receptor occupancy was way too low. We also looked at 50% receptor occupancy. That was also very low; 80% receptor occupancy based on our method was actually safe.

What about the activity data? These are very good for hazard identification, so having the complete package of activity data such as the ones that I've listed will give us an idea of what's the potential for this product to cause CRS, for instance, but when it comes to using the activity data for first in human dose selection, it was quite challenging. First of all, not all sponsors provided the EC50 data even when it could be obtained and it could be provided. We didn't have the EC50s for all INDs and when the sponsors provided the EC50 data, it was very selective. Let's say they did 4 or 5 activity assays but they only provided the EC50 for one of them. For 2 of the INDs, however, the sponsor provided the EC50 data for all activity assays conducted, and the range of EC50s for all of them.

This blue line that you see is actually from 1 of these 2 INDs where we have the mean EC50 for each assay conducted, and these are the mean values. The highest mean divided by the lowest mean is 10,000, and if you include the ranges for each of these, for each mean, there is a range, you get closer to something around 50,000 or 100,000. The range is really wide. Now, the question is which one is the most relevant EC50 for first in human dose selection? We don't know. We don't know, the tendency is to usually pick the lowest one, and if you pick the lowest one for the most conservative approach, then, the dose could be in a very low range such as nanogram dose, a homeopathic dose.

This is just for better visualize what I was saying. This is a sponsor. This is IND, another IND where they actually reported the mean EC50s for all studies. This is a mean EC50 for one study, another study, another one, another one, another one, and goes on; 20, 30, so if you take the mean EC50s, the range is about 10-fold, and if you take into account the range of all of them, it's over 1000-fold difference. Imagine if sponsors conduct more of these. There used to be a time that I remember there was 5 cytokines in the cytokine release assays. Now we have more and more cytokines. I don't even know these cytokines.

When should we stop? Also, let's say, for T-cell activations, how many markers do we want to look at? Some look at CD25, some look at CD69, and so on. It keeps being added. The more activity studies you have, the more EC50 data that you have. The KDs were usually in a much tighter range compared to EC 50s.

This is, again, to show that when a sponsor has conducted a side-by-side comparison of their product to other products in development or approved products, they were able to start at a dose that was not really a MABEL dose, so in this case, 2 checkpoint inhibitors. The first in human dose is actually at saturation when we used our equations on slide 6.

I think we can skip this. Frequency of administration for the antibodies. About 80% of them was every 2 week administration or less frequently. None of the antibodies that we looked at so far resulted in cytokine storm, and remember that cytokine storm was seen after a single dose administration of TGN1412, and we've gone very high on doses. We are either above saturation or, for one of the products, are still in a phase 1 dose escalation. They're at 90% receptor occupancy and I'll show you the data.

With all of them, however, we did have IRRs and CRS with all of them. That's very common, as expected. They've been manageable and all protocols had measures to take care of them, to monitor, to treat, and there were either meds, other treatments, but also some sponsors when they saw CRS, they amended the protocol to change the duration of infusion. This is how they actually managed and they were able to dose escalate. There might be a higher incidence of CRS with antibodies that have increased ADCC and I'll show you in another slide.

We collected the data to see whether sponsors that propose rigorous treatment for CRS were able to start at a higher dose and we did not see any relationship. This was mainly those that propose to do it prophylactically versus after occurrence of the first event, and there was no relationship. Dosing was staggered and I already talked about treatment for CRS. There was high variability in first in human clinical trial designs. Some were single patient, some were 3 plus 3, some intrapatient escalation via dose escalation increments were 3 to 10-fold in between cohorts. I'm getting into the actual examples.

This is a checkpoint inhibitor. This is an intrapatient dose escalation design. The start dose is at 10% receptor occupancy. They are escalating 3-fold, 3-fold, so half logs increments up to the 70% receptor occupancy. At this point, they're switching to a 3 plus 3 design and the intrapatient dose escalation took about a year from their initial ID submission to their desired dose of 900 milligram, the MAD, maximum administrate dose, was 3 years, and 900 milligram is already above saturation.

This is a checkpoint stimulator given every 3 weeks. The start dose is at 6% receptor occupancy. The dose increment 10-fold, 10-fold, and then switches to 3-fold or half logs. Initially, there are one patients in each cohorts and then it switches to a 3 plus 3 design at a little bit above saturation. The question is are patients at 6 microgram or 60 microgram dose levels benefiting from these doses? Can we potentially just observe them for a week? If there is no CRS and if there is no API related toxicities, can we escalate them to a higher dose on week 2? These are just open questions.

This is a checkpoint stimulator. The start dose is at 2%, receptor occupancy 6 microgram dose. They have gone all the way to the 240 milligrams, so they've already saturated single dose administration, and it took them 5 years from the time of IND submission to the dose of 240 milligram was 5 years. Again, the question is: are patients at these dose levels benefiting? This is just to show you another design and it's a 3 plus 3 right at the get-go, and they've gone all the way to the gram doses with acceptable toxicities.

What are our recommendations? We don't have any recommendation at this time for CD3 by specifics. We are collecting data, so hopefully, maybe in a year or so, we can have some recommendations for you for antibodies. If you are using activity data to select the first in human doses, we highly encourage you to optimize your doses because you saw how wide the range for EC50s could be. If you do not optimize, you could be at the higher range and right now, we don't know what will be the first in human dose if we use one of the higher EC50s because we have only two INDs with all the EC50s.

We still don't know what is the best way for dose selection for immuno-oncology products, but at least just because we look at receptor occupancy and we seem comfortable with 20 to 50% occupancy or even higher than that, up to 80% occupancy, at least whatever dose that you're proposing perhaps you can check your dose against what we've done just to get the sense of where you are, where you fall.

These are, again, open for discussion. Let's say you've done your method, you've selected the dose, and you compare it to what we've done, and you fall below 50% occupancy. Can we allow intrapatient dose escalation for these patients so to avoid sub-therapeutic doses? What if you fall between 50 to 75% receptor occupancy? Can we say one patient only and not do a 3 plus 3 at that point, maybe at a little bit higher dose do a 3 plus 3 design? If we want to do a 3 plus 3 design, what should be that place? At what point do we start a 3 plus 3 design? This is, again, open for negotiation, just something to think about.

Do we really need to do a half log dose increment right at the get-go or if the doses are very low, less than 50% occupancy, can we allow a little bit higher than that, maybe 5-fold and 10-fold? Some sponsors have done it with no difficulties. Is it okay if the sponsor do a side-by-side comparison of their product to other products in development and start at a little bit higher dose and not necessarily at a MABEL dose?

Here's an example. This is my last slide and these on top, these are products with first in human doses at less than 10% up to 50% receptor occupancy. These are the doses that have been given, highest human doses that have been given except for this one, which is the recommended human dose, and these are the ratios of highest human dose to the first in human dose and you could see that it could be very large. You see that, for some of them, it's up to 5 years from the time of IND submission to the time of declaring that dose.

One thing I wanted to mention here. I put these 2 CD40 antibodies just because, sometimes, we get INDs where the sponsors say, "I want to start at 30 milligram dose because there is another CD40 antibody in development and they're at 60 milligram dose, so I'm going to just start at 30 milligrams or 20 milligrams. I should be safe." These 2 antibodies are both IGG1 against CD40. One of them has no modification in the FC domain. This one has no fucose, and we heard that earlier. It can have increased ADCC activity.

For this antibody, they saw the first grade 4 IRR and CRS at a dose of 60 milligrams, whereas for this one, they saw a grade 4 CRS at a dose of 3.6 milligrams. You might say, "Perhaps this one has higher affinity for CD40. Actually, this one had less affinity for CD40. Why is this now tolerated less? Could it be, perhaps, because of increased ADCC activity? That's a possibility. Just because somebody else has another product against the same target, that doesn't mean the doses that would be tolerated for your antibody is the same dose. These are higher at saturation for the first in human dose and some are based on xenograft studies and some are based on other products.

I would like to thank the working group, Ramon Boody, John Layton, Michael Manning, and Emily Warren for helping me with this project.

Pasi Janne: Thank you. I'm Pasi and I'm from Dana-Farber Cancer Institute. I'm going to talk about the clinical development of osimertinib, an EGFR immune lung cancer and I hope it'll dovetail with what Darren talked about also previously in the preclinical development.

Just to remind you again, osimertinib is a mutant selective EGFR inhibitor, so as opposed to have equal potency against EGFR activating mutations and EGFR wild type, there's a differential effect of being more potent against the mutant, both the EGFR activating mutant as well as the EGFR T790M mutation compared to EGFR wild type, and the hope from all the preclinical studies, of course, was that this would create a wider therapeutic window and that we wouldn't see the EGFR mediated toxicities like we do with agents like erlotinib and afatinib.

As Darren mentioned in his preclinical talk, really, the dose that was thought to be efficacious or the starting dose thought to be efficacious based on the preclinical data was in the 20 milligrams, and this is the design of the phase 1 first [inaudible 00:22:42] clinical trial. This design incorporates a couple different unique features.

First of all, we knew exactly the population of patients to go into, and that was EGFR immune patients who had failed a prior EGFR inhibitor because we're testing a new EGFR inhibitor against the common resistance mechanism. Second, this was not a 3 plus 3 design. It was called a rolling 6 design and evaluated multiple different cohorts of patients with different doses of osimertinib in a trial that was conducted both in Asia, in multiple parts in Asia, as well as in the US. Of course, EGFR immune lung cancer is a problem around the world.

Secondly, if there's any activity that was observed in any of these cohorts after completion in the escalation cohorts, it allowed opening of expansion cohorts for patients who are then centrally tested for the common EGFR resistance mutation, T790M mutation, and were enrolled into cohorts of patients that have the mutation and those that did not, hence allowing in the same trial, then, to compare the preliminary activity in patients with the common resistance mutation versus those that did not. This was just the patients on the trial, just to highlight the point. The main point was that this was a trial that was, again, conducted around the world, had a large representation of Asian patients, both Japanese patients as well as non-Japanese Asian patients, and patients from Europe and the United States.

When we looked at the doses ... in the trial, an MTD was not defined and there were no DLTs identified in the trial, and so it went all the way from 20 to 240 milligrams. What you can see is that in each of these cohorts ... in this initial portion of the trial, there are 283 patients enrolled and you can see that, in each of the dosing cohorts, there was a large of patients which allowed us to then both evaluate the toxicity or the side effects as well as the efficacy.

What you can see here, if you look at the AEs across the different doses, you can see that between the 80 and 160 milligram dose, there's a little bit of an inflection in the AEs themselves. If you look at them in more detail and look at EGFR related AEs or the types of toxicities that we normally think about with EGFR inhibitors like erlonitib and afatinib which include things like rash or diarrhea, you can see, again, that they start to increase, although most of them are less than grade 3, but they start to increase between the 80 and 160 milligram dose, and that becomes an important feature as we think about what doses to take forward.

If you look at the efficacy, so I mentioned this, there was not an MTD defined and no DLTs. The efficacy was, again, evaluated in a large number of patients. These are the patients that now have the centrally tested T790M. You can see that the responses were actually quite similar across all of the patient populations based on the dose here, so responses ranging between 50 to 66%, you can see here, and they're color-coded based on the doses. If you looked in the patients that did not have the T790M mutation but were still resistant to conventional EGFR inhibitors, there was some activity. It was much lower than in the T790M positive patients and, again, didn't significantly vary between the different doses that were tested.

If you look at progression free survival in the early part of the phase 1 trial, as was hoped and anticipated, it was longer for the patients that have the common T790M mutation compared to the patients that did not have a T790M mutation as measured in this centrally tested patient population.

One of the other things about the trial, again, as it was conducted across different populations, we're able to look at the PK in patient populations. This was the different dosing levels. You can see the PK in the single doses, a dose proportional. After multiple dosing, this is at cycle 2 day 1. This is a steady state PK. It takes about

15 days to reach steady state and the drug has a half-life of about somewhere between 48 to 50 hours.

Again, we looked at the PK across or the exposure across the different ethnic population so here are Japanese patients, non-Japanese Asian patients, and patients from both Caucasian and non-Asian patients from around the world and, again, the exposure appeared to be similar across different ethnic subgroups, suggesting that there's likely to be ... or as we were thinking about the dosing decisions, one dose should be appropriate for all patient populations, and, again, the trial allowed us to do all of this at the same time.

Now, based on all those findings, the 80 milligram dose was picked as I mentioned. It was an efficacious dose. It was the inflection point before you started to see toxicity or more EGFR related toxicities and the trial subsequently had an extension cohort based on that 80 milligram dose in a separate phase 2 clinical trial that was also conducted also at the 80 milligram cohort, and so these make up the phase 2 components of the study. This is now an updated data from the phase 1 which, at the 80 milligram dose, has a 70% response rate, and these are the pooled phase 2 analyses which have a 66% response rate. You can see here very similar looking waterfall plots for these two cohorts and these were the PFSs just presented recently. PFS of 9.7 months from the phase 1 portion at the 80 milligram dose and 11 months from the pooled phase 2 portion of these trials.

Now, these initial studies led to the accelerated regulatory approval of osimertinib within a very short period of time. Initially approved in the United States in November of 2015 and within a few months in the EU as well as Japan. In fact, turned out to be one of the fastest approvals in oncology drug history. The first patient was entered in March of 2013 and the first regulatory approval in the US was in November of 2015 partly because we knew exactly the patient population to go into. The preclinical data suggested that it would be efficacious in T790M positive patients and the trial allowed us to, in a very rapid way, to ask and evaluate that question.

Subsequently, additional trials of completed enrollment, a phase 3 trial in patients that have T790M compared to chemotherapy and a frontline trial in patients compared to existing EGFR inhibitors, in this case, gefitinib or erlotinib.

Just to show you a couple of other examples of where the agent has been evaluated, because I think it gets into some of the issues that we've talked about here as well. Again, as part of the phase 1 trial, there was a first line cohort of patients where, actually, 2 different doses were evaluated, the 80 milligram dose and a higher 160 milligram dose, and this was to gain information about how would osimertinib perform in an EGFR TKI naïve patient population, so if you can overcome the most common mechanism of resistance, is it perhaps better to prevent it from happening in the first place? This was to get some information on this patient population.

In this patient population, these are just the adverse events from that patient population. An 80 milligram, again, similar to what we had seen in the previously treated patient population as well as in the 160 milligrams, except the one notable difference here is that, again, as we knew from before, the 160 milligram had somewhat higher incidence of AEs, and you can see here, somewhat higher incidence of dose reductions in this cohort compared to the 80 milligram cohort.

These are the responses in the untreated patient population. Overall response rate of about 77%; 67% for the 80 milligram and 87% for the 160 milligram, but, really, overlapping PFS curves here with the combined PFS of about 19.3 months. Again, suggesting although not proving that this is, potentially, a strategy that may be better when you start with this EGFR inhibitor that can overcome the most common mechanism resistance as the initial therapy. Obviously, as I mentioned, there's a phase 3 clinical trial that's completed enrollment, which we'll evaluate that formally to see if that, indeed, is the case.

One of the other activities that I learned about during the development of osimertinib was the penetration into the brain. This is a real problem for or brain metastasis and leptomeningeal disease, a real problem for many of our lung cancer patients as we develop better systemic therapies, we're seeing more and more disease and more and more relapses in the brain. It's a real problem in the [ALC 00:32:12] patient population and there are agents that now can effectively penetrate the CNS and the CSF, but until we started to learn about osimertinib being able to do that, we didn't really have those same kinds of therapies for the EGFR immune patients.

Additional preclinical studies were done where these were carbon labelled and given to cynomolgus monkeys and you can see here the uptake of osimertinib in the brain here. This is in comparison to ruxolitinib which doesn't have significant uptake, and here also gefintinib, which also doesn't have significant uptake in the brain. In a different kind of preclinical model, this in a xenograft model where an intercarotid injection was performed of EGFR immune cell lines to generate brain metastasis. You can see here the mice were then treated with various doses of osimertinib here and you can see regression of brain metastasis.

In fact, in the clinical trial, several patients had ... even in the phase 1 clinical trial, although it wasn't formally meant to evaluate brain metastasis, there were patients that had either received prior radiation that were allowed to go on the trial for their brain metastasis or had asymptomatic brain metastasis, and you can see activity here, for example, in these 2 instances and here in both patients seen here.

Now, more recently, another study has been performed. This was just recently presented at ASCO last week. Looking at the impact in leptomeningeal disease, so brain metastasis and leptomeningeal disease, again, are problems for patients that do develop CNS disease from lung cancer and, in many cases, leptomeningeal disease is really an access problem of drugs being able to penetrate the CSF and when there's very, very limited studies that have been performed on looking at, for

example, drug concentrations of gefitinib or erlotinib in the CSF and they're vanishingly small compared to what they are in the systemic circulation.

This trial called the Bloom Study specifically looked at patients that had leptomeningeal carcinomatosis and evaluated a higher dose of osimertinib. Here, this dose was picked with the idea that it would have a higher, perhaps higher CNS penetration although it wasn't ... this was not comparing doses. It was picked as a dose there to evaluate whether there is activity in patients who have leptomeningeal disease here and all of the patients here had to have EGFR immune lung cancer and positive CSF cytology and confirmation, obviously, leptomeningeal disease.

This was a limited study, but I think quite informative of 21 patients who had leptomeningeal disease. They all have had prior EGFR TKIs and about half of them had also had prior whole brain radiotherapy. What was interesting is that, of the patients treated, there were patients that were in 2 different categories. Patients that, at baseline, had a normal neurologic function or, at baseline, had some evidence of neurologic dysfunction as a result of their leptomeningeal disease, and a number of those patients, they had improvement in their neurologic dysfunction here. Here are the normal patients. Here are the patients with abnormal function at baseline. Some of them had improvement and there were definitely responses seen here as well as some clearing of CSF cytology in patients with leptomeningeal disease.

One of the challenges here is that there are no response criteria. There's no resist criteria that's even established of how you evaluate an effective agent for patients with leptomeningeal disease and hence we're trying to capture here different aspects of that.

One of the nice things that was done in this trial was PD in the CSF. We don't do so well on thoracic malignancy clinical trials of PD or do PD so well in patients that have systemic disease and the fact that PD was done in the CSF was remarkable. This was looking at CFDNA in the CSF over time in patients treated, and you can see here a very nice decrement in some of these patients in their CFD and a burden in the CSF as a function of treatment suggesting that the drug is getting into space and modulating, having its intended effect by seeing the reduction here.

This just shows you the duration of treatment that there are a number of patients that had significant durations of treatment here for 7 of them who had been on treatment for greater than that 9 months and a number of patients who had their CSF cytology clear. A nice example of potential use for an agent or additional benefits potential of an agent that wasn't necessarily appreciated upfront, but I think also fills in the need in the clinical need for patients who develop leptomeningeal disease as well as brain metastasis.

Just to summarize some of the highlights here for osimertinib, the clinical portion of it ... effective in patients with EGFR T790M. As I mentioned, there are multiple

effective doses from 20 to 240 milligrams that were evaluated in this extensive phase 1 program across ethnic populations. The 80 milligram dose was developed further and there, as I mentioned in the trial, the MTD was not established and there were no DLTs as a wide therapeutic window, and, again, effective in brain metastasis. There is additional ongoing development trials that have completed and an example when I showed you with development in patients specifically with leptomeningeal disease.

I forgot my acknowledgement slides but I, obviously, want to acknowledge all the investigators and patients and families and the collaboration with AstraZeneca with many of the members as part of this whole program. Thanks.

Ying Lu: My name is Ying Lu and I really appreciate the meeting organizer invite us to present some of our work and my information is in the handout, but I want also mention the work I present is mainly work of my colleague [Zi Lai 00:38:46] who's in China right now. He would rather be here and he's co-director for the Center of Innovative Study Design at Stanford and the papers, the joint work with [Lai Bachov 00:39:03] and, also, [Narushi Han 00:39:05] at Stanford.

The traditional way this is different from previous one ... we don't have data and we don't have example, but we could talk about methodologies. We approach, in the early phase of drug design, have 2 phases; phase 1 is usually understand [inaudible 00:39:23] the safety dose and, in the cytotoxics case, we use MTD but, in general, you want have some constraints about dose, so you identify under certain conditions. Keep in mind, in the phase 1 case, we actually don't know the true dose often, the target dose, so it's point of estimate. Then, the phase 2 part typically want to get signal for the efficacy and the typical way is to do the hypothesis testing. The goal is that if we have the right dose, then we want say if the drug is not active compared to the historical level also have acceptable activity, therefore let's go the next phase of development.

Here, keep in mind that EDA here in the phase 2, hopefully, is the true dose with the constraints, but in the practical way, that phase 2 never really linked with phase 1, so phase 2 take the fact that the phase 1 dose is true and you move on, and maybe later stage you find the dose is too high and that time, it's kind of late and you see the patient withdraw from treatment. That's a current practice.

We have the workshop here basically. People all know we not only, in early phase 1, look for safety dose, but also get early FS signal and there's some new patient methods being well developed and well known and people are using, but ready to review of a method known for the frequent disapproach, so the goal of this talk is really trying to introduce a very general approach called sequential generalized like the ratio test that can be used to evaluate both safety and efficacy in the combined phase 1 and 2 trials. The paper actually been published in 2014 statistic in medicine by Bachov Lai and Narushi Han.

I give some brief introduction about sequential test and then show some example

from the paper and some take home message. Just go back to a little explanation what is like the ratio, and we know the likelihood is basically ... you have a random variable and they come from a family of distribution. If you have continuous variable, then the chance observe is represent by density function. If you have discrete case, you can represent as probability. The data may come from a family of distributions which we don't know what the [theta 00:42:11] is the distribution premise which is unknown, and [inaudible 00:42:16] interesting to compare 2 possible choices of where the data from a family of parameter theta 1 versus data from family of parameter of theta 2, then you do the likelihood ratio, basically, how likely your data is from a theta 1 distribution versus theta 2 distribution, and this called likelihood ratio for one observation.

When you have a lot of data, you can cumulatively calculate a likelihood ratio to observe the set of this data and that is well known called the Neyman-Pearson Iemma if you've not learned the basic statistics, the likelihood ratio test is most uniformly most powerful test. Now, what's different here with sequential likelihood ratio. It's based on a ration test, but it's not collecting data all at once. Instead, you have sequential steps, you collect data, you look the likelihood the ratio. At the time your likelihood ratio is strong enough to differentiate 2 groups, then you stop. Therefore, keep in mind the sample size itself is a random variable, depends on what you observe.

That's the difference between Neyman-Pearson and the water sequential statistics. As water in develop for the weapon system in 1945 and he conjected that will be the optimal design and, later, he proved. There are a lot of extension from that now. In the phase 3 trial, it is well known that people do sequential clinical trials for [inaudible 00:43:46] analysis both fatality and efficacy, so it's not very full in concept. What relative less is it not being used in the early phase.

If you want to learn more about sequential experiment design and sequential likelihood method, there is a new book published by Bachov Lai and Nashi which have a lot of things they mentioned in this presentation.

Now, we move on for the phase 1, 2 trials. Here, we're basically in the proposed method is, first of all, we need a joint model for the toxicity and the efficacy, and model can be any generalized. It can be parametric continuous version or even discrete dose groups. The combined approach start from the typical phase 1. You can use any kind of phase 1 approach because hypothesis testing phase is really in the second stage for efficacy, so the first part, what you do is select a basic potential dose and [inaudible 00:44:51] constraints. Then, in the stage 1, you add additional patient at your recommended dose and combine phase 1 data together as your first intro stage group.

Over there, you look for a ration test. If you can reject [inaudible 00:45:11] hypothesis that the treatment is, indeed, effective, then you can stop at using that recommended dose. Otherwise, you will continue to add additional group patients. When you look for the additional group patients, you assign the patients to refine

the dose, because you now have more patient than what you had in the phase 1. Every time you can update your constraints about condition of your safety profile and then you continue to lock like a function.

The way that propose here instead of test, we test for the dose [ee-da inaudible 00:45:48] which is satisfy your constraints and then compare their efficacy probability. The method I'd recommend from that paper using the habitual [peet-uhl 00:45:58] boundary which have 3 parameters. One is if you want reject no hypothesis, you, first of all, look for your response rate, and if the response rate, indeed, is above the P0. Secondly, the likelihood ratio of your data is higher than the boundary and, here, the likelihood ratio means you observe maximum likelihood estimator as your true parameter versus the no hypothesis which no affects it. If the data is more in favor of your no likelihood ratio test and you feel comfortable to reject it. Otherwise, if your observed response rate is below your acceptable P1 and, also, your likelihood ratio is much lower than the P1, then you can say, "Okay. I will say the drug doesn't work, so accept no hypothesis."

Otherwise, in between, you can continue add patient data until last group, and you only make one decision there. If you reject, if not reject, you accept no hypothesis. The procedure is mechanical and pretty straight forward. There is implementation in that paper hall to select the BP tutor and the C to maintain your type 1 and type 2 error rates.

Here, it just shows simulation to help understanding why we want to do it. First of all is we look for the current practice and, of course, we can only select one model. If, say, we do the dose escalation at phase 1 with [ee-wok 00:47:39] model and then we, in the phase 2 part, we're using assignment 2-stage approach, of course, these for the cytotoxic models, but there is a measure, again, the test is on the efficacy side and the dose is really as a constraints, no hypothesis testing. It's only updated. I just want mention that.

This first example is, actually, non-case which, basically, we say at a MTD level, the treatment has empirical which is .1 instead of .25. In the Simon hypothesis testing, you can set different type of type 1 error rates, so we put .01%, 2%, all the way 5% and with 80% power. The truth here is we are in the no hypothesis setting and the joint model, basically, is just regression model which we use to generate the data. Here's take home message. If you look for the red line, it is nominal. It is planned. Type 1 error rate and solid black line is what you actually add up the percentage of your reject no hypothesis, you say the drug works at MTD level and partial [inaudible 00:49:02] because, in the typical way, when we do the Simon test, we don't go back to see if the drug recommendation is right so maybe sometimes you had over than the MTD, therefore, you get chance to reject no hypothesis, because your effect rate is truly above that. That's where misleading that you go to the phase 3 studies possible.

The second example is really compare the typical approach versus the one that's sequential like a ration test. The basic model are the same, the difference in the

second stage that instead of we do Simon stage, we do 5-group sequential study. We divide for the Simon sample size for the 3 patients 10, 10, 10 plus 3 in the last group. The first group, actually, we have 24 plus 10 patients in the sequential analysis. The cutoff value was selected calibrate so that we have the same type 1 error rate in order to fair compare the power and the sample size. The red light representing new test, and you can see it's higher power than the Simon's approach. Also, it's less type 1 error when, actually, it's less than the .01 efficacy rate. Also, the sample size is much less.

Now, I want to say the sample size here compares is not complete fair because Simon's test, actually, only stop for futility and when its efficacy ... Simon's philosophy is you want go more data to get more information to help you design phase 3. Still, even given the sample size, we would not increase the sample size. That's the point, take home message. There are also, as an advantage still, you can look into including the percentage of patient being receive overdoses much less and, also, residual roomy square error is much smaller and overdose is reduced. Overall response rate is higher, so there advantage showing here that actually work well. The second example is in the papers on discreet, so the first example 2 was on the continuous dose, which is not practical. In the most case, you have only finite selection of doses.

The simulation's 3 examples shows when you have finite dose selection, the basic take home message it still works, so I just skip that. Now, we also say this is generalized model and we actually working on looking for some other options. One possible way is to look for the joint efficacy modeling for optimal biological dose. Here, we use the [inaudible 00:52:06] paper and the trinomial model, so the idea here is the toxicity is a logistic regression model, but the response rate, depending on the result toxicity, you have another model for conditioning on logistic regression model for response, and the goal here is trying to find the dose that not only minimize toxicity but, also, maximize the response without toxicity, so it's optimal, so that's what that definition is.

You have the likelihood function and we can write out and do the similar [inaudible 00:52:43]. Another extension I want mention because we had talk a lot about exposure model, I just want bring it up from Lai Chan and, also, confounding, this was a paper that in '89 in the mathematical sciences and Lai Chan named that distribution called unnamed distribution, and the idea here is that you have expose at a time ... if this X is dose, this [pus-aye 00:53:11] is acceptance absorb rate, and, so, if we have a instantaneous dose, it's a multiplication of the time. Then, the body discharge, so clear the drug was exponential rate, and, so, that's what Chan's basic idea. Using this one, one can pretty much model any kind of exposure between times. Say, if the drug administrated during S1, S2 time window and at the time of T, you can calculate remaining dose the current concentration would be.

Also, you can calculate cumulative dose by integration of this concentration. What Chan mentioned that basically is the tumor incidence rates in his paper has a ratio, is a function of cumulative dose at a time of T, so you can have multiple exposure

of a time as long as the remaining dose is there, you can calculate incidence rate. He fit that model for cancer time, for HIV, AIDS on time and very well done. We want borrow that idea for the toxicity and efficacy. Basically, we borrow that idea for exposure, then we can model the hazard ratio for toxicity event as a function of, say, dose concentration or area under our C-curves. You can model the response rate based on the area in the cumulative dose or also based on whatever the driving effect of exposure. We can build likelihood function and we can, basically, move to the likelihood ratio test.

The event [inaudible 00:54:52] Chan and confounding model is it's flexible, so you can address drug sequencing, you can put different time there. Once the model fit, you can also evaluate the late toxicity if the model's right, of course. The disadvantage, of course, we don't have any data to fit and see if that actually biologically work, and that's something that with work with UCSF group in trying to learn, so it's work in progress. The take home message is that standard practice have inflate high point error rate when you only look for phase 2 and don't look for phase 1. The artificial barrier between phase 1 and phase 2 in our new design has been removed because your phase 2 phase will still keep on updating the constraint of the dose.

Group sequential methods, very general method, and it's improved both average sample size and the power for phase 1, 2 combination. There's also our package available free for download. If you want test the method, please feel free to contact. You can read that paper, which give the link for the software or if you send e-mail to me, I will be happy to provide. Thank you. The last one, thankful Lai for the comment.

Chao Liu: Good afternoon, everyone. I want to make sure: is everyone able to hear me? Okay. Good afternoon. First of all, I would like to say it's a great honor to make this presentation for this section. My name is Chao Liu and a [inaudible 00:56:46] developer at FDA. Today, I would like to share some experience on [inaudible 00:56:51] response analysis to assist in both the marketing dose automization trial. Especially, when dose adjustment, the rate is high due to drug toxicity, how [inaudible 00:57:02] response could be conducted to handle this issue. I will talk about a recent FDA case reveal on lenvatinib for renal cell carcinoma.

In this submission, high dose adjustment rate was observed due to the drug toxicity and a post-marketing requirement was issued to optimize the dose. To figure out the most promising dose of regimen to study, to explore response analysis was performed, to evaluate various dosing regimens. A noble method, dose adjustment integrated exposure response analysis. We also call it DAEIR was supplied to handle the drug toxicity cost dose adjustment. With this type of technique, different and studied dosing regimens could be evaluated in a more realistic settings through assimilation.

It is the first time we performed this type of analysis. We hope our work is able to provide some [inaudible 00:58:09] information to the community in terms of how

to deal with the similar issues in the future. Before starting the presentation, I would like to make a standard disclaimer that the views in this presentation is my personal opinion.

In the first part of the presentation, I will give a brief background on post-marketing dose optimization trial and explore a response analysis. I will then talk about the case review on lenvatinib for renal cell carcinoma where DAEIR analysis was able to support the post-marketing dose [inaudible 00:58:46] trial design. A short summary will be given at the end of the talk.

When pre-marketing trials suggested a studied dose might not be optimized, post-marketing dose optimization trial is conducted to maximize the drug effectiveness and minimize its toxicity. In oncology, one major cost to conduct such trial is that the drug safety is a concern where dose adjustment could [inaudible 00:59:18] to mitigate the drug toxicity. In addition, in terms of the post-marketing trials, one of the challenges is that how to adequately use the limited patient resource to find out the best dosing regiments. In many cases, only one alternative dosing regimens can be studied due to practical issues.

Therefore, one primary issues for the trial design is that to find out which dosing regimen to study. It relies on maximizing the learning from the available data and is where modern assimilation could be helpful. To explore a response analysis, quantify the relationship between the drug exporter and the efficacy or safety. The ER model can be used to design the optimal trial through trial of assimilations. In our case's review on lenvatinib, the ER analysis was performed to choose the most promising dose and regimen in the future post-marketing dose optimization trial.

Lenvatinib is a receptor tyrosine-kinase inhibitor that was originally approved as a first line therapy to treat the differentiated thyroid cancer. Recently, lenvatinib was approved for the treatment of advanced or metastatic renal cell carcinoma as a second line in combination with everolimus. The approved dose is 18 milligrams lenvatinib plus 5 milligram everolimus QD. In the [inaudible 01:01:04] trial, patient in the lenvatinib-everolimus combination arm, which is the blue line in the Kaplan-Meier plot for progression 3 survival showed significant improvement in PFS as compared with the lenvatinib and everolimus monotherapy. Overall survival was also observed, however, 89% patient in the combination army experienced a dose reduction or interruption due to the drug toxicity.

[inaudible 01:01:38] one major concern about the current dose, thus, FDA issued a post-marketing requirement to optimize the dose via conducting a post marketing study, but which dosing regimen to study? To answer this question and spur a response modeling assimilation was performed to find out the most promising candidate of those new regimens. The evaluation of and study the dosing regimen via ER analysis is composed of 2 major steps: modeling and assimilation. The modeling step quantified the ER relationship for drug efficacy and safety based on available data. Next, trial assimilation will be performed to evaluate different dosing regimens.

For each dosing regimen, the dose exposure profile will be firstly predicted and efficacy and the safety profile is then simulated based on the generated exposure profile and the ER relationship. For lenvatinib, 3 types of dosing regimens were evaluated. Lower lenvatinib doses, dose holidays, and the lower doses plus op titrations. One major challenge to our analysis is how to handle the dose adjustment treatment by the drug toxicity. In the next few slides, I will discuss the traditional approach of ER analysis in oncology and why it is not able to deal with this issue. In addition, I will talk about how the new method, DAEIR, handles this problem.

In the traditional ER analysis, exposure was assumed to be constant over time throughout most part of the trial. When dose does not change, when constant exposure matrix, for instance, steady state AUC can be used to represent the exposure level for each subject. The ER relationship is then estimated by associating this constant exposure matrix to a efficacy endpoint, say, progression free survival. In the ER relationship, it can be estimated based on such association.

In addition, in assimilation step, drug exposure is predicted based on a predefined dose level, assuming dose does not change in response to the post-treatment events. For example, adverse events. When dose adjustment is minimal, the traditional approach is okay to use. However, dose adjustments cost by intolerable adverse reactions raises challenges for the traditional approach of the ER analysis.

The occurrence of intolerable AEs results in dose interruption reduction and, in one subject, makes appearance multiple dose reductions during the trial. In this case, the exposure will be driven by the altered dose and, thus, varying over time. A so-called constant exposure cannot be found in this case. Can we just calculate the drug exposure based on dose intensity or average dose and use this so-called average exposure to do the ER analysis? No. I'm going to show you why it's not going to work. Assume the progression free survival is used as the efficacy endpoint. For a subject who progresses soon, the duration of the treatment would be short. He or she may have no chance to experience multiple dose reductions and, thus, still remains in the higher dose level. Average exposure could be higher.

On the other hand, a subject who progresses later stay longer in the trial and, thus, has higher chance to experience more dose reductions and winds up with a lower average exposure. With the traditional approach, the ER relationship would appear to be flat or even reversed and, thus, give a biased result. Finally, because the dose is dynamically driven, in response to an intolerable adverse reactions in the trial, the dose exposure safety interaction should be considered in assimilation. The traditional ER method cannot incorporate this part.

To address these challenges, a [inaudible 01:06:53] dose adjustment integrated exposure response analysis was used. Instead of choosing a constant exposure matrix over time, time varying exposure was implied. In this case of lenvatinib, it is the AUC at each time interval. in addition, as the dose exposure changes during the

trial, the drug, in fact, would also be varying over time. To capture this, the longitudinal tumor size was used to represent the drug efficacy. In terms of the ER for safety, occurrence of intolerable AEs was associated with concurrent exposure, that is AUC at the corresponding time interval. Finally, to assimilation step to address the adverse reactions cost dose adjustments during the trial. DAEIR dynamically generates the dosing history to incorporate the dose exposure safety interaction.

I will give a brief summary about the DAEIR analysis for lenvatinib. In the modeling step, ER relationship for the time course of tumor size as well as adverse reactions leading to dose adjustment was estimated according to the available data. In the trial assimilation step, a dosing history was dynamically generated based on a ER relationship for dose [inaudible 01:08:36] adverse reactions. Next, the time course of tumor sizes was then predicted on the basis of the generated exposure profile and exposure efficacy model.

The exposure efficacy relationship between lenvatinib-everolimus exposure and a time course of tumor size was explored using data from the registration trial. In this model, the tumor growth rate equals the natural growth rate minus suppression, in fact, from everolimus and lenvatinib. According to this model, higher lenvatinib exposure is associated with a greater tumor suppression. Meanwhile, the ER relationship for dose altering AEs was also estimated. AEs leading to dose adjustment was treated as one repeated events. Through communication with FDA, a longitudinal [inaudible 01:09:37] model for adverse events leading to dose adjustment were developed by the sponsor. This model will be used to simulate a dosing history so as to incorporate the dose adjustment cost by intolerable adverse reactions.

These slides represent an example of simulated dosing history where a starting dose of lenvatinib was selected as 12 milligram. The X-axis is the time after treatment and the Y-axis is the percentage of each dose level over time. The dose level is represented by the responding color. As shown in the table, the dose adjustment was pre-planned and executed upon occurrence of intolerable AEs. In this example, by sixth month, about 32% of the patient still remains at their original dose level, which means they did not experience any dose altering AEs in the first 6 months.

Based on the generating dosing history, the tumor dynamics was simulated to evaluate the efficacy at each dosing regimens. At each plot, the X-axis is the time of the treatment up to 1 year, and the Y-axis is the relative tumor size as compared with the baseline. Each colored curve is the simulated population mean values of a tumor dynamics and/or each dosing regimens. The dosing regimen of 18 milligram lenvatinib plus 5 milligram everolimus served as the control. We firstly evaluate if simply lowering the lenvatinib dose would provide comparable efficacy. Dosing regimens of 14 milligram, 12 milligram, or 10 milligram lenvatinib plus 5 milligram everolimus was evaluated, and none of them was able to provide a same magnitude of tumor suppression as compared with the control.

Next, adding drug holidays to the original dosing regimens was also evaluated. In this setting, patient who took 18 milligram lenvatinib and a 5 milligram of everolimus for 2 consecutive weeks were allowed to have 1-week dose break. However, the simulation results show that this type of regimen could also result in a compromised efficacy. In the end, we found that the implementation of an op titration option might work. In this scenario, I patient was allowed to be op titrated to a higher lenvatinib dose level every 4 weeks if he or she did not experience any intolerable adverse reactions before. The dose cap of lenvatinib was set to be 18 milligram.

When op titration option was provided, lower lenvatinib dose could provide comparable tumor suppressions as compared with the control. Therefore, in terms of the regulatory decisions, a post-marketing requirement was issued to optimize the dose. According to the simulation results, 14 milligram of lenvatinib plus 5 milligram of everolimus were selected as alternative dosing regimens in the study. The take home message is that the dose adjustment integrated ER analysis, the DAIER, can be used to optimize the dosing regimen and, also, to incorporate a dose adjustment caused by drug toxicity.

In the analysis we presented, we have some assumptions. For example, we assumed a same ER relationship for all dose altering AEs. In addition, only target lesions were included in exposure efficacy analysis. Finally, I would like to thank everyone who involved in this review work and also with collaboration from the sponsor. Thank you.

Matthew Guo: Hi, good afternoon. First of all, I would like to thank AECR and FDA that invite me. [inaudible 01:14:39]. I'm here to share dose optimization study as a post-market requirement. The E7080 is a code name for the lenvatinib which just mentioned by Lui. It is [inaudible 01:15:03] an FGFR targeted, that agency. The material that included in this presentation are the result of works from many colleagues in [ayz-aye 01:15:17] and, also, the collaborations with FDA and, of course, the patients, so I would like to acknowledge that up front.

Here, I'm going to provide real world example to share stories how a drug going through the dose finding and up to the FDA's approval and then result the post-market requirement. The study that leaded to that FDA approval is phase 1B and 2 study of lenvatinib alone and also in combination with the everolimus in the subject with RCC. In the phase 1B, it is traditional phase 3 plus 3 study design, trying to find a MTD and followed by small, the extension study to confirm that recommended phase 2 study. In the phase 2 study, that actually is randomized with 3 arms and with a combination and lenvatinib, that alone, and everolimus alone.

Again, the phase 31B is a traditional 3 plus 3 study and the starting cohort is that everolimus 5 milligram plus lenvatinib 12 milligram. The everolimus is ... they approved the treatment in RCC and the dose is 10 milligram daily. Also, we have the approved label for the lenvatinib in the DTC with 24 milligram daily [inaudible

01:17:35]. In order to test the combinations, we started with both single drug after the half of those approve the dose. It is escalated into the cohort 2 which was everolimus 5 milligram daily and the lenvatinib 18 milligram daily. Then, further escalated into the everolimus 5 milligram daily and the lenvatinib 24 milligram daily. Then, result the 2 DLT, so we considered ... that expanded the cohort 2 to the 11 subjects that recommend the phase 2 dose move forward into the phase 2 study.

As I said that, it is randomized, open label, 3 arm study with lenvatinib 18 milligram combined with the everolimus 5 milligram daily. Then, the second arm is lenvatinib 24 milligram daily and the third arm is everolimus 10 milligram daily. The primary endpoint is PFS by the investigator and the enrollment started of the March of '12 and ended at the July 2013. Then, the data cutoff for the primary analysis is June of 2014 when the [nan-dee 01:19:13] PFS event [inaudible 01:19:16] event is occurred. Then, these are the results briefly mentioned by Dr. Liu, this a [inaudible 01:19:28] for the PFS that, according to the SAP. As you can see that this line is a combo arm. This is the everolimus arm and the medium PFS is almost tripled compared to the everolimus and a result of the P value of 0.005 and with the hazard ratio of 0.4. The number of subject in each arm is 51, 52, and 50.

Then, also, at the time, we also look at the overall survival and result has a ratio of .5, so this is a combo arm. This is the everolimus arm. With the confidence interval, the [inaudible 01:20:22] 95 confidence interval barely over 1 and a result P value of 0.062 basically demonstrate that there's a trend in the overall benefits overall survival as well. These are the total ... the exposure, as you can see that, for the median duration of the treatment, the combo arm is almost the double of the everolimus. This is certainly correlated with the efficacy PFS, that endpoint. It is also, somehow, that reflected that a subject is able to tolerate this treatment as well.

This is overview of the AE. As you can see that the grade 3 AE is comparable to the lenvatinib single arm. Basically, indicated as combo does not increase the overall toxicities with the lenvatinib single arm, but it does have more toxicities compared to the everolimus. The other [inaudible 01:21:53] that SAE that, again, indicates that it does not increase very much compared to the single arm, but it does have slightly increase in terms of the everolimus. [inaudible 01:22:16] can say is that the dose modification. Even though it does not increase dramatically compared to the single arm but, somehow, it is more compared to the everolimus. We have to keep in mind that, for this specific combo in the study, the design for the 205, coded as 205, we intentionally to start with the high dose and in the prodigal had a great details and how to manage the dose in the anticipation that certain AEs will happen and the dose will be reduced. We tried to minimize a potential of a tumor drug resistance.

Also, when you talk about the drug withdraw, it's almost double compared to the everolimus but, also, we have to keep in mind that the exposure is also almost double. Certainly, that long exposure could lead to the longer drug withdraw due to

... even though they are the competing risk with the PFS endpoint. On the July of 27, 2015, FDA granted breakthrough therapy for this indication and on the November of 2015, [ayz-aye 01:24:02] filed supplemental NDA and it was classified as priority review and on May of 2016, middle May of 2016, FDA approved the indication of lenvatinib in combination with the everolimus is approved for treatment of patients with [inaudible 01:24:23] following [inaudible 01:24:26] antiangiogenic therapy.

As mentioned by Chao Liu and, also, at the same time, and issued a post-market requirement that also had a detailed timelines to the proceed with this post-marketing requirement. Out of concern that there's a lot of those reductions and drug was draw as well. I'm going to [inaudible 01:25:04] details regarding the dose optimization study design. We had many conference was conducted between FDA modeling simulation group and subsequently with the bio stats group and between [ayz-aye 01:25:23]. Then, both FDA and [ayz-aye 01:25:29] conducted multiple scenarios regarding dose schema. Then, we also consult with external the expert regarding the basic design because we anticipated there could be a possibility we have multiple dose that we have to look into then, we have to that implement the phasing strategy how to quickly that eliminate the dose.

This is one of the slide from communication with the FDA and indicated it also would show in one of Dr. Liu's slides that as well. With the 18 milligram, which was this pinkish line. Without those escalate up and compared to this 14 milligram which is agreeing with starting those with the 14 milligram but with the option of those escalate titrate up. There is a potential [inaudible 01:26:44] that efficacy wise that [inaudible 01:26:48] a tumor reduction could be even better than the 18 milligram as starting dose. With this information armed, so we proceeded with a dose optimization study as post-market the requirement. With the double-blind and the randomized double-blind with the one arm with the 18 milligram plus 18 milligram lenvatinib plus 5 milligram everolimus. Then, the second arm is the starting dose with the lenvatinib 14 milligram with the everolimus 5 milligram but with the option that if it doesn't observe those intolerable grade 2 or grade 3 AE, then this 14 milligram arm subject could titrate up to 18.

Then, also, both arm will allow dose reduction follow those schemas similar to the study 205. This is a title of the prodigal and the objective is trying to look into this 2 starting dose, trying to achieve the objective of the similar efficacy, but better safety profile with the lower starting dose. The primary endpoint ... actually, there's 2. One is the objective response rate at the 24 weeks by independent review, and the second primary endpoint is actually the safety endpoint which is intolerable grade 2 and grade 3 and above by the week 24. Then, we also look at the other endpoints, as well.

We implement a [inaudible 01:29:04] trial and look at the efficacy and, also, test the superiority for the safety endpoint. We assume that 37% response rate for the 18 milligram and 45, the response rate, for the 14 milligram starting dose and with the [inaudible 01:29:36] margin of 0.76, so it requires 306 subjects which is 1:1

ratio between to arm to have 80% of the statistical power and one side that off of 0.05. Also, with this sample size, we are able to detect 15% a rate of reduction with 2 side off of 0.05 and with the 80% of power. We also implemented 2 interim analysis. One is at the 150, one is at 200 total subject once they complete 24 weeks follow-up or the discontinued [inaudible 01:30:26].

In terms in [inaudible 01:30:30] margin, it was a set up 0.76, so when, after the end of study, if we test the 40 milligram starting dose compared with the 18 milligram starting dose, if the lower limit 90% confidence interval is above 0.76 in terms of ratio wise, then we will that declare the 14 milligram has similar efficacy compared to 18 milligram starting dose. In order to help the physician understand what this mean, we also converted this model in terms of rate difference. For example, if the 18 milligram, the response rate is 50%, then, somehow, the 14 milligram this difference minus 18 milligram if the 90% lower limit is above negative 0.07, then we will declare it is [inaudible 01:31:38].

Also, for the interim analysis, here's a boundary. For the noninferiority, we use [inaudible 01:31:50] a frame, strong boundary. Then, for the futility, we also improvement the futility, we used the integrated nonbinding stopping boundary. Here's both the cumulative [inaudible 01:32:08] and also this is cumulative beta [inaudible 01:32:12]. This is current study design. Thank you.

Eric Rubin: We have about 15 minutes, I think, so I think we'll have an abbreviate panel discussion if I could ask the speakers and panelists to come up on the stage. Again, for time purposes, I won't introduce the panelists who haven't spoken. Please, if you have a comment, introduce yourself. While people are assembling, I'll start. We've heard, in this session, an interesting approach varying from case studies to a proposed statistical approach and then to the FDA's interesting data on what's happened with MABEL. I guess one of the questions I'd ask, and this is a theme from earlier today is: should we recommend that there always be a randomized component to dose finding?

I noted that it's varied, I think, amongst speakers and among the case studies some have done. Some have done it, some have not. I sit next to Mark Ritain who's now left who is a strong proponent of including randomized dose findings. I think it may be good to understand, particularly, for biologics where it's not even feasible, often, to target an MTD because you're never going to have it. To get to a recommended dose, should we recommend that there be randomized dose findings? I'll just open that up to the panel if someone wants to take that on or anyone in the audience can also chime in as well.

Matthew Guo: Can I comment on [inaudible 01:34:08]?

Eric Rubin: Sure, yes. Yeah.

Matthew Guo: I think, certainly, for the dose finding, if we have the randomized multiple arm, that would be very ideal, but I think the challenge is at how do you define the optimal

dose? When you look at the toxicity sometime, you can look at within the cycle, but when you look at the efficacy, you really needed to either have a very good biomark, you are able to ... that observed you in that shorter period of time. Otherwise, if you don't get the PFS and the OS, it will take a very long time.

Eric Rubin: Tumor shrinkage is one that people have used, right? I think that's fairly early biomarker that can be used for efficacy, right?

Matthew Guo: Right, certainly, but I think, at the one point, at least, the people even do not believe that PFS actually predict the overall survival, which is ultimate clinical endpoint. That was a challenge at the time. If you're going back, that tumor shrinkage might be even propose more challenges. If you look at the tumor shrinkage, then you have to also look at the remain shrinked for how long as well.

Eric Rubin: Yeah. I guess we could talk about endpoints, but it goes back to ... in the absence of randomized studies, I guess the concern is that we're going to be doing those in the post-marketing setting. Is it better to do some of that up front rather than take the risk of having to do it in a post-marketing study? You can certainly include elements such as OS or PFS in those. They're later endpoints ... anyone else want to comment on that topic? Lei?

Lei Nie: Yeah. This is Lei Nie. I am statistician from the FDA. I guess, for the comparison, based on testing or dose response the modeling approach, I would prefer modeling approach. Not necessary comparison. In terms of randomization, I believe you have more dose randomization certainly helps. Thanks.

Serban Ghiorghiu: Maybe to build ... [inaudible 01:36:40] Serban Ghiorghiu from AstraZeneca. I'm the global clinical lead build [inaudible 01:36:44]. I don't think that randomization will be a solution for everything rather than ... and will not be a size fit all. In some cases, will be randomization, but in some cases, as Matthew said, tumor shrinkage. Not all agents will be associated with tumor shrinkage. In other cases, thinking of going back to some of the ideas, there are so much data between us. Maybe we can look at something more [inaudible 01:37:13] controls because we'll have the patients, we can do those comparisons.

NA. Rahman: I'm Atico Rahman. I'm from the FDA. I think from the FDA perspective, we would always support our unpromoted randomized dose finding trials because it has got a number of advantages. It gives, even, estimates of effects based on exposure or dose that you're testing. Also, it could be a carrot for accelerated approval if the data is robust enough, and would give you a better understanding of what post-marketing or final registration of your trial could be. I could give you a better estimate of what does you could take forward, what regimen you can take forward. I think from [inaudible 01:38:13] perspective, we would always support randomized dose finding trial.

Kelvin Dickerson: I think from a purely external lens, I would say the researcher should be approaching the work with the intention to disprove the hypothesis and

randomization is a way that you can aid in that because it prevents any bias being subliminally introduced through selection. I think my only other observation on that would be I think there's a certain ... we've seen a very wide range in sample sizes, and I think there should be some minimum requirements for statistically relevant sample sizes.

Eric Rubin: Yeah, thanks. I don't think this really came up in any examples, but I've certainly seen many of this recently, especially in these days of combinations with immunotherapy where ... I think the last time I counted, the field was up to over 500 combination studies. I know for [Murk 01:39:14], we're at about the 170 mark. Many of these involve combinations with either [sen-er-eh-tive 01:39:22] care or other novel agents, and it surprises me that oftentimes the decision to move forward into an expansion cohort or subsequent effort is based upon 6 patients or 10 patients or something. It's a little bit like the 3 plus 3 that we focused on last year. I think that the idea that you can make a decision on dose with 6 patients, I think, is unbelievable, but still seems to be quite prevalent. Again, I wonder if that's something we can focus on it coming out of this meeting as something not to do.

Ying Lu: I think whether you do the randomized study or not is the aim, what you really want achieve, right? The goal is that safety that you want really make sure the 2 arms have comparable safety or efficacy. I mean, with all that, it's hard to understand whether you want require or not, but given this morning's talk, given such a large sample size, it sounds like a rational to consider more than just exploratory and, really, generally, something to answer [inaudible 01:40:36] 1,000 patients phase 1 study which really opened eye for me to see such a large phase 1 study.

Eric Rubin: Yeah, well, I think as pointed out, I wouldn't call them phase 1s. These are phase 1 through 3. It started off that way.

Ying Lu: The difference in the phase 3, you have predefined goals and I'm not quite sure, here, that [crosstalk 01:41:02].

Eric Rubin: Yeah, they can't, but you can define it there, too. In our particular example, we predefined. We had 8 amendments. Each of them had a predefined hypothesis. It's basically just, again, compressing all of your development into one study. Although it's, again, as has been discussed in other scenarios ... not for the faint of heart and not to be recommended as a general way to do drug development.

The other one I wanted to talk about was Dr. Saber. I found Dr. Saber's presentation quite interesting. Thank you for sharing that data, and was struck by some of the ranges and doses and the time it would take to get to a maximum administered dose or even a recommended dose. What I wondered was the fold increases between doses, especially if we're starting at a very low dose. I don't know if we've a lot about that today, but it's come up at times. It seems, to me, at least that that is somewhat empiric and I wonder what people think about ... if you're starting at a very low dose, what would be appropriate fold increases with

subsequent doses?

Haleh Saber: Today's presentation is really to open up the room for discussion and to hear from everybody what they think. We've seen half log increments at very low doses. We've seen 10-fold increases and I think for one of the INDs, actually, between the cohort 1 and 2 was 50-fold. There is no standard and they're all pretty much at the same low doses and it seems that it worked fine but the doses are very low, so should we really ... traditionally, for biologic, we used to see half log increases. I know that some people still like the half log. Maybe they do not know that, actually, more than that could be done. We would actually like to hear from the audience, from you, from the clinicians what they think should be the increments.

Eric Rubin: Well I think I'll just say you're [inaudible 01:43:23] true because I think I know, and sometimes my Merck colleagues, they come in conservative and they say, "Well, the FDA will never let us go beyond this conservative approach." It's actually very comforting to hear you say that can be a consideration. I know, again, from my own perspective. One of the more difficult consents I have is a patient with one of these trials where you're concerned that they're getting a subtherapeutic dose that if there was a way to get them to, as you commented, by interpatient dose escalation is something that's a little closer to efficacy, assuming they're not having side effects. That, to me, makes it an easier discussion with the patient.

Haleh Saber: Right. I think it's easier to agree on intrapatient doses, at least my personal opinion, than to agree on 10-fold versus 20-fold versus 5-fold increase-

Eric Rubin: At a cohort level.

Haleh Saber: Yes, so if the doses are very low, let's say you are 10%, 20%, 30% receptor occupancy and you dose the patient, you have 1-week observation, no drug related toxicity, can we allow that patient to go to the higher dose level and then continue with that? Is it something we should be thinking? Should allow this?

Eric Rubin: Yeah, thanks. Other panelists? Audience?

Kevin Dykstra: Kevin Dykstra from qPharmetra. One of the talks we heard earlier today had a very nice demonstration of a PK simulation where they looked at the dose intervals and were looking for adequate separation. Rather than trying to agree on a fixed one, it might be a good idea to try to simulate the PK that you expect and look for a reasonable separation in the doses. As a consulting person, one of the things that I always try to discourage my clients with is the 10% increment in the dose based on the single incidence of an adverse effect and so forth. That would be one perspective.

Eric Rubin: I was thinking even more that more aggressive, again, around 10, 20-fold which I think, historically, we tend to shy away from those, but if you're at a very tiny dose in the beginning-

Kevin Dykstra: Whatever you have the guts to do.

Eric Rubin: Yeah. Right.

Lei Nie: I think this is a good idea. Rather than fix the dose, fold changing dose calibrate by, for example, from O to dose chance. For example from 20% to 40% O change. What is the resulting fold change in dose? That could be an idea. I particularly like the escalation for the intrasubject, then move onto intersubject. That's a very good topic for statistician to evaluate, I think.

Eric Rubin: Yes, thank you.

Male: Can I have a follow-up question for doctor Saber? I was wondering if you, when you looked at the data for immuno-oncology agents, if you looked at it differently for agonist versus antagonist? Because some might be the ... the prevailing notion is for antagonist, you might need a higher level of receptor occupancy to actually achieve efficacy and for an agonist, you might get away with a smaller receptor occupancy.

Haleh Saber: I think, if I understand the question correctly, you're asking about activities, not safety? Whether the agonist versus antagonist, at a different receptor occupancy, we see activity.

Male: Yeah, then presumably to the extent that safety is a exacerbation of the [inaudible 01:47:04] activity, they may be related.

Haleh Saber: In terms of activity, we try to look at activity in phase 1 clinical trials, but, unfortunately, we couldn't get very good data, so what we did was that ... are we seeing activity, cellular activity, when we compare it to in vitro activity data? All the first in human doses are supposed to give certain level of cellular activity. That said, that doesn't mean antitumor activity, and we couldn't find any biomarkers of activity at doses tested to be able to see what type of activity is seen in humans in a phase 1. Sometimes, its sponsors mention, and the investigators were sure, there was activity and they just left it at that, not mentioning at what dose level they saw that activity. What is the definition of activity? We couldn't obtain that data, actually.

Sumithra Mandrekar: Can I make a couple of comments on the design aspects? One, I want to say that intrapatient dose escalation is always exciting, but when you start thinking about, those data are no longer independent. They're coming from the same patient. Then, you could also have interpatient dose escalation within the same design. Then, I think we have to be careful about how we use this data to dose escalate the next patient, new patient, so that can be a challenge because you have correlated data versus independent data.

The second thing about how do you pick the different doses, I think you had a great idea but, also, maybe, in identifying discreet dose levels and trying to step through

them, it is possible ... this is just something I was thinking as the talks were going on. You identify a region that is low-dose level, mid-dose level region or something and then a high-dose level, not really specifying. It's 1 milligram per kilogram or 2, 3 like that. Maybe something in the lower dose range, something in the middle dose range, and something in the high dose range, and start your dose escalation study using a model based design to help you figure out which region you want to hone in on. Depending on how the safety data come out, you may end up refining the doses between the low and the medium versus low and the high, so I think there are some other new ways to think about how to pick your dose space when you design such studies.

Eric Rubin: Thanks. One more question, then maybe we'll go to closing comments from Dr. Kim and McKee after that. Please.

Female: My question is for Dr. Saber. During the IND reviews, did you also focus on monotherapy versus combination? The reason I'm asking is the step [inaudible 01:50:01] receptor occupancy that is required for monotherapy may be very different for combinations.

Haleh Saber: That's a good question. We only focused on monotherapy, so everything I presented today was on monotherapy. Going back to the question on activity, I don't know if you remember the second scenario I presented for activity. Who asked the question on the activity? Okay. The second example I presented for activity, for that one, for instance, the very low dose, which was 6% receptor occupancy had activity at a cellular level, but antitumor activity was at a much higher dose, half mg per kg. One mg per kg is a dose that actually is the highest dose right now. ND investigators were sure the sponsored mentioned that there was activity, that's it.

Eric Rubin: I'd like, again, thank the panel for an abbreviated discussion and the speakers as well. Thank you.

Amy McKee: Okay. This is going to be extremely brief. I think what this day has shown is that we might need a part 3 in a year. Part 1 was 2 days, part 2 was 1 day, and I don't know how long we need part 3 to be, but I think there were more questions brought up at the end of the day that still need to be answered both by the science, by the modeling, by the clinic. I want to thank AACR, I want to thank all of our speakers and panelists, I want to thank all of our attendees, both in person and online. I think we have some very interesting discussions to come in the future, and I'll just leave you with ... we have 2 more workshops coming up, one on cardio-oncology, and one on immuno-oncology. They're both in September, so look out for those on FGA websites, AACR websites, and some of our other sponsors. Thank you very much and safe travels to wherever you are going.