U.S. FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH (CDER)

CENTER FOR BIOLOGIC EVALUATION AND RESEARCH (CBER)

Public Meeting:

Promoting the Use of Complex Innovative Designs in Clinical Trials

Tuesday, March 20, 2018

8:35 a.m.

FDA White Oak Campus

10903 New Hampshire Avenue

Silver Spring, Maryland 20903

Reported by: KeVon Congo,

Capital Reporting Company

A P P E A R A N C E S

Deborah Ashby, PhD
Imperial College London
Julie Beitz, MD
Office of Drug Evaluation III, CDER, FDA
Robyn Bent
Office of Biostatistics, CDER, FDA
Scott Berry, PhD
Berry Consultants, LLC
Frank Bretz, PhD
Novartis
Aloka Chakravarty, PhD
Office of Biostatistics, CDER, FDA
Shein-Chung Chow, PhD
Office of Biostatistics, CDER, FDA
Ivan Chan, PhD
AbbVie Inc.
Scott Emerson, MD, PhD
University of Washington
Steven Goodman, MD, PhD
Stanford University School of Medicine

A P P E A R A N C E S

Frank Harrell, Jr., PhD
Vanderbilt University School of Medicine
Telba Irony, PhD
Office of Biostatistics and Epidemiology, CBER, FDA
Laura Lee Johnson, PhD
Office of Biostatistics, CDER, FDA
Stefanie Kraus
Office of Regulatory Counsel, CDER, FDA
Lisa LaVange, PhD
University of North Carolina at Chapel Hill
J. Jack Lee, DDS, MD
MD Anderson Cancer Center
Gracie Lieberman
Genentech
Gregory Levin, PhD
Office of Biostatistics, CDER, FDA
Roger Lewis, MD, PhD
University of California, Los Angeles;
Harbor-UCLA Medical Center

A P P E A R A N C E S

Olga Marchenko, PhD
Bayer
Cyrus Mehta, PhD
Cytel Corporation; Harvard University
Will Meurer, MD
University of Michigan Health System
Thomas Permutt, PhD
Office of Biostatistics, CDER, FDA
Dionne Price, PhD
Office of Biostatistics, CDER, FDA
Karen Price, PhD
Eli Lilly and Company
John Scott, PhD
Division of Biostatistics, CBER, FDA
Rajeshwari Sridhara, PhD
Office of Biostatistics, FDA
Joseph Toerner, MD
Division of Anti-Infective Products, CDER, FDA
Z. John Zhong, PhD
Biogen

C O N T E N T S

C O N T E N T S

C O N T E N T S

P R O C E E D I N G S

DR. PRICE: Good morning. We're going to go ahead and get started, realizing that there are many attendees that are trying to get through the lobby security right now. But we'll go ahead and get started, realizing that in the afternoon we'll have to keep a check on the weather.

So we'll -- I'll ask Dr. Aloka Chakravarty, who's the acting director of the Office of Biostatistics in CDER to begin with welcoming remarks.

WELCOME, OPENING REMARKS AND INTRODUCTIONS

DR. CHAKRAVARTY: Good morning, everyone. I'm Aloka Chakravarty, acting director of the Office of Biostatistics, CDER, FDA.

Welcome to the FDA public workshop, public meeting on promoting the use of complex innovative designs in clinical trials, which is being convened by the Center for Drug Evaluation and Research, and the Center for Biologics Evaluation and Research.

We're pleased to be joined for these discussions by leading experts across government, academia, industry, and care delivery for a productive

exchange on the issues on-hand.

Our purpose today is to facilitate discussion and information sharing about the use of complex innovative designs in drug development and regulatory decision-making.

This meeting fulfills obligations under both 21st Century Cures Act and PDUFA-VI to convene a public meeting to discuss various complex adaptive, Bayesian and other novel clinical trial designs with particular focus on clinical trial designs for which simulations are necessary to valid operating characteristics.

We have seen examples of innovative designs at various stages of development. But its use can be improved by consistent acceptance in regulatory decision process of such designs and clarity may be needed to -- how to proceed with such a design.

Today's meeting is meant to be the beginning of an ongoing effort to discuss and explore the use of complex innovative designs in drugs and biologics. And it will be a chance for the leading experts to discuss their experiences with these techniques and to provide input as FDA develops a pilot program for complex

innovative designs.

Throughout the day, we encourage meeting participants to consider opportunities for increased collaboration, both across industry and with agency to support these efforts.

In a few moments, we will hear from the speakers about their thoughts on complex innovative designs and where they'll see potential for additional work.

We will be specifically talking about four topics: a session on complex adaptive clinical trial designs, a session focused on other innovative designs including external or historical control subjects, Bayesian designs and master protocols, a session looking at clinical trial simulation for confirmatory trial design and planning and a session where stakeholders share their thoughts about the upcoming pilot program for complex innovative design.

Before we get started, a few housekeeping notes. As you'll note in the agenda, each session will begin with 15-minute presentations, followed by panel discussion. We also have time set aside for broader

discussion with audience participants.

For those in attendance, we have microphones setup in the room, in the aisle, for you to use during the question-and-answer during the day. Formal comments can be sent to the open docket. The link is listed on the agenda.

I want to remind everyone that this is a public meeting, and the event is being broadcast online. So everything you say will be part of the record. For those in the room today, if you need to purchase a lunch, please be sure to order and pay for it at the Sodexho kiosk before the end of the first break.

Finally, a reminder that although this meeting is being convened by the FDA, it's not a federal advisory committee. The meeting will be a success if there is a robust discussion of ideas and open discussion.

So with that, I would like to thank all the panelists and the audience for participation, and we welcome and look forward to an active discussion. Thank you.

DR. PRICE: Thank you, Dr. Chakravarty. Before we being our first presentation, I would like to thank all of our panelists for traveling near and far to be with us today. We look forward to a very interesting and robust discussion.

I would ask that you each introduce yourself, giving your name and affiliation, and we have further information at the desk about your actual biography. So I'll start with Ivan.

DR. CHAN: Hi. My name's Ivan Chan. I work at AbbVie.

DR. LEWIS: My name is Roger Lewis. I'm the chair of emergency medicine at Harbor-UCLA Medical Center, and the senior medical scientist at Barry Consultants.

DR. ZHONG: My name is John Zhong and I'm from Biogen.

DR. LEE: Hi. My name is Jack Lee. I'm from University of Texas, MD Anderson Cancer Center. I'm a professor of biostatistics and associate vice president of -- president of quantitative sciences.

DR. MARCHENKO: I'm Olga Marchenko from Bayer.

DR. PRICE: Hi. I'm Karen Price from Eli Lilly and Company.

DR. HARRELL: I'm Frank Harrell. I'm with the Office of Biostatistics, FDA, CDER and with Vanderbilt University, Department of Biostatistics.

DR. TOERNER: Hi. I'm Joe Toerner. I'm at FDA, CDER, in the Division of Anti-Infective Products.

DR. SCOTT: I'm John Scott. I'm in the Office of Biostatistics and Epidemiology in FDA, CBER.

MS. BENT: I'm Robyn Bent. I'm in the Office of Biostatistics CDER.

DR. PRICE: I am Dionne Price. I am the acting deputy director of the Office of Biostatistics, CDER.

DR. LEVIN: HI. Greg Levin, statistician. Office of Biostatistics, CDER.

DR. CHOW: This is Shein Chow from the Office of Biostatistics, CDER.

DR. LAVANGE: Good morning. I'm Lisa LaVange. I'm associate chair and professor in biostatistics at University of North Carolina, Chapel Hill.

DR. ASHBY: Hello. I'm Deborah Ashby from

Imperial College London, statistician.

DR. BRETZ: Frank Bretz, Novartis.

DR. EMERSON: Scott Emerson, professor emeritus, biostatistics, University of Washington, Seattle.

MS. LIEBERMAN: Good morning. I'm Gracie Lieberman from Genentech.

DR. BERRY: Scott Berry, biostatistician, Berry Consultants.

DR. MEHTA: Good morning. I'm Cyrus Mehta, president and cofounder of Cytel Corporation and adjunct professor of biostatistics at Harvard University.

DR. PRICE: And if I could ask my colleague, Lauren Sucher, to stand in the audience. She is our representative from the press office today. Thank you. And one minor correction: The bios are online. They're not outside on the table.

So without further ado, our first presenter, Dr. Greg Levin, will begin.

SESSION I: GENERAL CONSIDERATIONS FOR COMPLEX ADAPTIVE CLINICAL TRIAL DESIGNS TO SUPPORT THE EFFECTIVENESS AND

SAFETY OF DRUGS OR BIOLOGICS

PRESENTATION

DR. LEVIN: Hi. Good morning. I'm going to provide a brief presentation to introduce and hopefully stimulate a good discussion in session one.

Session one focuses on general considerations for complex adaptive designs. The primary focus of this session is on adaptive designs that will be used to support the effectiveness and safety of a drug or biologic, although we expect that some of the considerations discussed will also be useful for early phase exploratory trials.

I'm going to provide a brief overview of adaptive clinical trial designs, including a definition to outline the scope of today's discussion. And then I'm going to discuss some important considerations for adaptive designs that I expect will be discussed in the session this morning, and I'll conclude by briefly going through the questions that we're going to ask the panel to discuss this morning.

So our definition -- our working definition of an adaptive design for today's discussion is a clinical

trial design that allows for prospectively planned modifications to one or more aspects of the design, based on accumulating data from subjects in the study.

Importantly, there are two scenarios that are therefore not within the scope of today's discussion. The first is when there are unplanned changes based on comparative interim results. For example, dropping of a dose because of unexpected toxicity.

The second is when there is information from sources external to the study. For example, results from a different study of a different drug that might motivate changes to the ongoing study in the form of a protocol amendment.

These are important scenarios that come up that could warrant another discussion. But today's session is focusing on adaptive designs that are prospectively planned.

One way to classify adaptive designs is by the type of adaptation that is being made at an interim analysis. For example, there can be adaptations based on baseline characteristics, such as a covariate adaptive design that attempts to reduce imbalance

between treatment arms.

There can be adaptations based on pooled outcome data, often called blinded adaptations. For example, to modify the sample size at an interim analysis based on pooled blinded estimates of the variants or the event rate, or there can be adaptations based on comparative interim results, often called unblended adaptations, such as group sequential designs that allow stopping for efficacy or futility, or designs that allow adaptations to the sample size, to the patient population, to the treatment arms included in the trial, et cetera.

Another way to classify adaptive designs is according to whether there are adaptations to statistical aspects of the design, such as the sample size, or whether there are adaptations to scientific aspects of the design, such as the patient population, the treatment arms in the trial, the endpoints.

When there are adaptations to scientific aspects of the design, the primary estimand, i.e., the primary measure of drug effect that we are trying to estimate in the clinical trial, will change

accordingly.

One of the motivations for adaptation is that in some cases, an adaptive design can provide advantages in statistical efficiency over a non-adaptive design, such as a greater chance of detecting a drug effect at a given expected sample size.

In some cases an adaptive design can also provide ethical advantages. For example, the opportunity to stop a trial for futility or efficacy can help ensure that patients inside the trial are not exposed to unnecessary risks, and that patients outside the trial are provided promising therapeutic alternatives as soon as possible.

And in some cases, adaptive designs can also provide advantages in the understanding of drug effects, such as an improved estimation of the dose response relationship.

On the other hand, there are some limitations and challenges of adaptive designs. There are methodology challenges in ensuring the control of the chance of erroneous conclusions and ensuring the reliability of treatment effect estimates. There can

be added operational challenges in maintaining confidentiality to comparative interim results and ensuring trial integrity, and there can be potential challenges in interpretability and generalizability due to changes in the estimand of interest during the trial.

In some cases, adaptive designs can be quite complex in that they may include multiple types of adaptations such as an interim analysis that allows adaptations to both the treatment arms in the trial and the sample size that will be accrued before the next analysis.

They could include adaptations to scientific aspects of the design, and many often involve simulations to evaluate operating characteristics at the planning stage.

A couple examples of adaptive designs that have been carried out that have some complex adaptations involved include PREVAIL II, which was a trial to evaluate ZMapp for Ebolavirus disease, and included frequent interim analyses with decision rules based on Bayesian posterior probabilities and allowed

the opportunity to add experimental agents as treatment arms during the trial if they became available.

And another example is I-Spy 2, which was a Phase II trial to screen breast cancer treatments and included potential adaptations to the sample size, the randomization ratio, and the treatment arms.

This slide lists a number of important considerations for adaptive designs that I expect will be discussed in the panel discussion this morning. I'm going to go into these in a little more detail on the coming slides.

The first is the control of a chance of erroneous conclusions. We also have the extent of reliability of estimation of treatment effects, the extent of pre-specification of details of design, the maintenance of confidentiality to comparative interim results and the extent of documentation, both prior to and during the trial.

One important consideration is the extent to which the chance of erroneous conclusions is controlled in the trial, and this includes the control of incorrect conclusion -- the chance of incorrect

conclusions of safety or effectiveness, of incorrect conclusions of a lack of safety or effectiveness and of incorrect benefit-risk evaluations due to misleading estimates.

One important component of the evaluation of effectiveness is typically the test of a null hypothesis in a clinical trial. And it is well-known that the use of adaptations in a trial can inflate this type one error probability without appropriate use of adaptive testing methods that have been supported by theory or comprehensive simulation.

Another important consideration is the reliability of treatment effect estimates. The availability of accurate and precise estimates help facilitate a reliable benefit-risk evaluation and appropriate labeling and reporting of results to enable evidence-based medicine.

Adaptations can induce bias in estimates, and some methods have been developed to have more desirable properties. An important topic for today's discussion is the extent to which this bias should be evaluated and the extent to which methods, where available,

should be used for reporting.

Another important consideration is the extent of pre-specification. This can vary and could include things such as the anticipated number and timing of interim analyses, the type of adaptation, the statistical methods for interim and final analyses and the algorithm governing the adaptation decision.

Possible motivation for pre-specification include that it facilitates the use of appropriate inferential methods for many types of adaptations. It can help increase confidence that adaptations are not based on accumulating knowledge in an unplanned way, and it can help motivate careful planning and monitoring.

As I mentioned previously, the scope of today's discussion is on prospectively planned adaptive designs. But one important topic within that scope is the extent to which that should all be flushed out and documented at the design stage.

Another important consideration is the preservation of trial integrity. It is recommended in ICH E9 guidance that access to comparative interim

results in all trials is limited to individuals independent of personnel conducting or managing the trial. And there are some added logistical challenges in maintaining confidentiality to interim results when you have an adaptive design.

Special considerations include whether there's use of a dedicated adaptation committee or whether the DMC is instead tasked with implementing the adaptive design, the use of confidentiality agreements, firewalls, data access plans and whether steps are taken to minimize knowledge that can be inferred through the adaptive decisions.

The documentation for an adaptive design can also be more comprehensive than is typical, and may include things such as the rationale for the design, the evaluation of important operating characteristics, the adaptation, monitoring and data access plans and, in some cases, a comprehensive simulation report.

There are also a number of other considerations that I have listed here, but am not going to go through in too much detail, but that may come up during the discussion this morning: the use of

simulations in planning and the role of Bayesian adaptive designs are parts of sessions that will come later today and will generate a lot of discussion.

There are some special considerations for adaptations in time-to-event settings, such as the role of nuisance parameters, such as the enrollment rate or the censoring distribution in trial planning.

There are special considerations for adaptations based on potential surrogate or intermediate endpoints, such as the modeling of the relationship and the assumptions about the relationship between the intermediate endpoint and the clinical outcome of interest.

And the final consideration that I'll mention is the importance of the evaluation of safety in adaptive clinical trials. For example, there may be a minimum number of patients or a minimum duration of follow-up that is expected for a reliable safety evaluation, and this can certainly impact the nature and timing of interim analyses that will be appropriate in an adaptive design.

With that, I'm going to briefly read through

the three questions that we're going to ask the panel to discuss this morning. The first question is the following: What are the two to three most important principals for sharing the appropriate and effective use of complex adaptive designs.

Our second question for discussion is the following: Discuss the extent to which complex adaptive designs should be pre-specified. For example, discuss the importance of pre-specifications of the specific algorithm that will be used to determine adaptive decision-making.

And finally, bias and treatment effect estimation is currently less well studied that type one error probability control in the context of complex adaptive designs. How important is the evaluation of the properties of point and interval estimates? Should adjusted estimates be included in labeling and reporting of results?

That concludes my presentation. I'm really looking forward to just a great discussion this morning. I'm going to turn the microphone over to our two primary discussants for session one, who are going

to provide reaction to my presentation and any introductory comments to get the discussion started. So I think we'll start with Dr. Bretz, who has a slide here to present to you all.

DISCUSSION

DR. BRETZ: Good. Okay. Good morning, and thanks to the organizers for inviting me to this very important panel. And congratulations to Greg for this very comprehensive overview.

I would like just to add one perspective before we go into the actual discussions, with respect to the three questions that Greg had posed, and my comment is about the usability of adaptive clinical trials and that it really depends on the specific application.

And in order to illustrate my point, I wanted to bring one analogy. Since I'm based in Switzerland, I thought I'd introduce the Swiss Army knife. And think about you have a very well-defined task about cutting a piece of paper. And, if you'd please click once, then which of the tools you would like to use? And probably everybody would use a simple scissor

because it is the optimal tool for a very specific task, namely to cut the piece of paper.

Now if you'd click once more please, you may recognize this is a Swiss Army knife, which is a versatile tool that combines several individual functions within a single unit, one for every perceivable need. So these knives are often used in the general -- as a general phrase to -- as an analogy or as a metaphor for usefulness and adaptability.

And the point I'm trying to make is that with a Swiss Army knife, you can still cut a piece of paper. You see the little scissor toward the bottom right. However, the scissor is not optimal because it's small in size, but you can still do a reasonable job.

Now, if you think about doing some other task, you can still use the Swiss Army knife reasonably well, so that if you really don’t know what -- you want to do it in advance or you only have a rough idea, then you can do multiple things with this tool, and then the Swiss Army knife is probably a good tool to use.

However, you can also overdo the things, and if you click once more, you see the picture of a real

knife that you can actually buy. It has about 200 functions, one for any need that you can imagine, and it weighs about two pounds.

Now, imagine there is somewhere among these 200 functions, there is also a little knife. Now, imagine you really want to cut the piece of paper with this little knife. You can imagine that this becomes complex. It becomes really difficult.

So while this tool really looks very, very impressive and probably -- well, it's expensive, but still you would like to have it. Once you have it, I can tell you I tried to cut a piece of paper and it doesn't work very well. So you can spend a lot of money, but in the end, you just put it into the shelf.

So be careful. And that's my real point, is that in applying adaptive design, do please consider also simple tools, such as a scissor, as they're very appropriate in many cases. But definitely avoid the so-called giant Swiss Army knives in clinical trial practice, I guess. So thank you.

DR. EMERSON: So first, thanks for presenting the issues. There's a couple points that I just want

to make for emphasis on the things. I guess the -- I always put it under a category of it's very important as we talk about any clinical trial design is what's the burden of proof.

And very often I find the biggest mistake is trying to use one clinical trial to answer every question, when you could more efficiently answer all the different questions of safety, efficacy and effectivess, perhaps by focusing on some different trials.

I also like to always think about, you know, what do we need at the end and then, what we need at the end is we need an indication for the drug, which involves what's the disease, the definition of the disease that we're using, what's the population that we think we'll use it in, what's the exact treatment that we think we would have, which is a complete treatment regimen and then, what's the outcome we're looking for?

And any of those that change, you've really changed the indication. But at the end of the day, we also need to write a label. And what do we need on the label that the physicians can use this treatment,

recognizing that the clinical trial always has a comparator that is generally not the comparator that the physicians are considering.

That is to say, they're not always just considering this drug versus placebo. They're considering this drug versus another drug or another sort of treatment, whatever the standard of care is. And so, being able to write a label so that people can understand what's there is very important.

And then the last part, I am, you know, merely closet Bayesian. But it's the Bayesian questions that are the most important. We want the treatment that is approved to be one that we think there's a high probability that it works. And so, always thinking back to the fact that as we start on drug development, the vast majority of drugs that we think work don't.

And so, it's very important to remember that as we go through this process, confirmation is very, very important because the positive predictive value after a Phase II study is necessarily much less than one after a Phase III study.

And the problems that we have as we come to

adaptation, if you start off with a drug that truly works at the level that you think in your Phase III protocol, all Phase III studies would be positive, right? I mean if you really -- if you have 90 percent power, 90 percent of Phase III studies would work, and they don't. And that's just because the prior probability is less than a hundred percent.

And so, just making certain that we focus, as we talk about adaptation, of what of these adaptations are very, very appropriate at the early stage, and a huge improvement.

And I like the Swiss Army knife analogy. I often remark to my hiking companions that what if you really know that the only tool you need is a corkscrew. Do I approve you bringing just the corkscrew on the hike? And no, I want the Swiss Army knife. I want it to have a few more tools, but not too much.

And my point would be even at the confirmatory Phase III study, limited flexibility is still within the confirmatory aspect. But it's going to be very important to talk about what's the prior knowledge that you have that you're contributing the confirmation to.

DR. PRICE: So thank you both for those, for your reactions to the presentation. I will ask Dr. Meurer -- and I apologize if I'm mispronouncing your last name -- to please introduce yourself, just giving your affiliation. And if we have any other panelists that may have joined us, feel free to come up to the panel.

DR. MEURER: That was an excellent pronunciation. I'm William Meurer. I'm an associate professor of emergency medicine and neurology at the University of Michigan.

DR. PRICE: Thank you. And Dr. Goodman, as you are getting settled, if you'd like to introduce yourself as well?

DR. GOODMAN: Sure. Thanks. I've been here listening, so I missed it. Steve Goodman. I'm a professor of medicine and epidemiology at Stanford.

DISCUSSION

DR. PRICE: Thank you. So we are going to move into our first discussion. We plan to do this by -- this is not an AC. It's a scientific discussion.

But we're going to use a similar method. If you could just kind of raise your hand, Robyn and I will look, and we will acknowledge. We will keep a running list of those people that would like to answer a question.

So we're going to start now. Again, the first question is what are the two to three most important principles for ensuring the appropriate and effective use of complex adaptive designs. And I see Dr. Mehta.

DR. MEHTA: I think there are three really important issues that must be addressed in adaptive designs. The first I think most important is not to disturb the equipoise of the investigators because they -- as long as they feel comfortable randomizing patients, they will participate. But in an interactive design, changes are made in a midcourse. And so, there has to be a lot of care not to disturb equipoise.

My second important point is to keep the interim data very secure and auditable so that you know at the end of the trial, if there is to be a final analysis, it should be possible to actually document in an auditable matter what was the state of the data at the interim analysis because the data from the interim

analysis will be combined with the data from the second stage.

And the third point is perhaps, everyone would agree, that you must simulate the operating characteristics of the design.

DR. PRICE: Dr. Lewis?

DR. LEWIS: Thank you, Cyrus. I think one of the first comments, which has already been partially alluded to, is that we need to be very careful to match the adaptive design to the true threats to trial success, where success is defined by getting the right answer, whatever that turns out to be.

I consider a trial to have failed if, at the end of the day, you don’t have a clear and correct answer to the primary question that motivated that trial.

I think there's a tendency to go the other direction, which is try to pretend that the real questions are the ones for which you have a solution as opposed to letting the true threats to the success drive the selection of the design.

The second point I'd make is that the design

choices that we make, that affect the balance of various complex goals -- and I'm thinking of things like bias and variance tradeoffs, the dual purpose of controlling error rates, traditional type one/type two error rates versus bias correction -- really should be guided by what we anticipate to be the actual use of the result of the trial, either to drive regulatory decision-making or clinical decisions at the bedside.

And as a practicing clinician, it's my observation that clinicians tend to use the overall qualitative result of trials in making treatment decisions. They very rarely look in any detail at the precise estimate of the treatment effect and that, at least in that context, the overall error rate is a much more important in consideration than the accuracy or lack of bias in estimates of treatment effect.

And then, if I will, I'd just say in terms of the equipoise comment, which I wasn’t going to comment on until you brought it up, Alex London at Pittsburgh has written some really nice philosophical work on the concept of equipoise as it applies to adaptive trials that use response adaptive randomization. So I just

want to make sure that people know that there's a really good resource on that particular topic.

DR. LEE: Okay. I just want to add a little bit on what's been said. I think that we all want to make the correct -- quote, unquote, "correct" or "accurate" decision. And from FDA's point of view, FDA would like to approve safe and efficacious treatment.

And regarding the inference, I think as a statistician, we all know that they are two main different ways of making inference into the inference framework, you know, the frequentists and the Bayesian.

I'd like to say that in the past, these two approaches has been very competitive. They fight with each other. Currently, it's more competitive, okay? And we are here today to talk about complex, novel design. And again, you can look at from the Bayesian point of view or frequentist point of view.

And I think the future, I think it would be more collaborative in the sense that type one -- yes, type one/type two error rate are important, but not the only thing. We will also need to look at the posterior probability of how effective a treatment is, okay?

So I think from the inferential point of view, it's not just a traditional point estimate or confidence interval estimate, you know? Yes, that's important, but also we need to look at what's the probability of success in a sense, right?

And I want to say a little bit more on this. That is, I think, you know, Steve worked a lot -- Steve Goodman, on these compared two different philosophy or the framework.

But one really limited approach in frequentist approach is this non-significant hypothesis testing, you know? That's kind of a -- you know, it has its role, but it's a very limited role. So let me just point that out, and I think we can have more discussion.

But through all this, I think three things are very important. One is education. Okay. The second one is innovation. The third one is implementation, okay? So I think we really need to do better in all three aspects.

DR. PRICE: So we have Dr. Ashby, followed by Dr. Marchenko, followed by Dr. Goodman.

DR. ASHBY: Thank you. Firstly, thank you for very clear presentations. I mean, to me the single most important thing is that you need to think really carefully what you're trying to do. Now, that's true for any study. And the whole debate on estimands actually says to me that you have to think really clearly what the purpose of the trial is.

But when you begin to think adapting, you have to think why are you adapting because that drives the design. It drives all of your decisions. So just in a dose response trial, are you trying to hone in on the best dose to take forward to the next stage or are you trying to learn about the dose-response curve because that will, to some extent, influence what the best adaptation algorithm is.

Secondly, it is plan, plan, plan, and again that's true for all trials. But for adaptive studies, you're generally doing it because you want to be really efficient. You've got one shot at it. And so the simulation, the discussions, maybe the rehearsals, you know, the decision-making processes can't overemphasize it.

My third point is really more of a question because I think when we're flipping around between some things which are really kind of more appropriate to early stage studies and some which are more confirmatory studies. So the third question we've got about what you put in the labeling would be very appropriate if you're doing some adaptation in a confirmatory study.

Actually, it seems to be almost irrelevant if you're doing an adaptation of dose on a first-in-demand study with a limited endpoint. And I just wonder whether just passing out some of those general considerations might actually help clarify the debate.

DR. LEVIN: Yeah. This is Greg Levin. Yeah. Thanks for that point. I mean, I'd say that our primary focus, what we're hoping is that we primarily focus the discussion on trials that would be intended to support determination of safety and effectiveness, so at the confirmatory stage.

But we expect that many of those considerations, like getting the question right and planning also have a role in early phase exploratory

trials. But that is our primary focus, and that is why that very question is focusing, for example, on reporting results.

DR. ASHBY: Okay. That's very helpful.

DR. MARCHENKO: Okay. Yeah. I just wanted to add a little bit. I think I'm a big proponent of adaptive designs. And I think even if in the end we don’t use adaptive designs, they allow us to understand, or at least evaluate it, quantify some uncertainties and then definitely help with planning better designs.

But I do want to tell -- or at least remind everyone that, in the end, what we want to do, we want to improve patient care. So we shouldn't just think in terms of specific adaptive designs for specific trial. We need to think in terms of programs and go even further because I did have an experience previously when the program went very quickly through Phase I, II, III and got approved. But then payers did not actually want to pay for the drug.

And in the end, we didn’t improve patient care because we didn’t have enough information on specific

endpoint of interest, like overall survival. So we do need to think in terms of progress rather than individual trials.

DR. PRICE: If I could ask if you're not speaking, please turn your mic off. That may help with some of the projection of sound problems that we're having towards the back of the room. So if you're not speaking, if you could turn -- push the red button. Turn your mic off. Thank you.

DR. GOODMAN: So, oh, that sounds loud enough. I missed the introduction unfortunately. So I don't know if there's like a dividing line between complex and adaptive. I hope there is because there's simple adaptive, too. We should not have complex inalterably next to adaptive because current trials are adaptive, just having samples -- you know, stopping rules is adaptive. So there are all sorts of things that -- forms of adaptation, stopping for futility in one arm, that we do all the time.

So let's not always say complex adaptive. Complex is a choice and, as has already been made clear, many of these trials, particularly in the

confirmatory phase, don’t have to be that complex. We might adapt on one thing or two things, and usually we're adapting for one of two reasons, either ethics, or efficacy -- or efficiency.

And the efficiency is either that we don't experiment in areas where we don’t -- where the conclusion is clear, or we're borrowing strength internally from things that are mutually informative. We don’t want to ignore that,

So in the confirmatory stage, usually the dimensions of adaptation are much more limited. So let's say adaptive, and then only invoke complex -- and I love Dr. Bretz’s -- the Swiss knife analogy. That is a very appropriate.

On the issue of equipoise, we have to be pretty careful here. You know, there's individual investigator equipoise. And then, there's sort of population equipoise, which is really the most important thing. That was the innovation that Friedman brought to the concept, which was it's disagreement in the treating community and not just in the individual investigator. It's epistemic. It's ethically

important.

And we always have to think about whether the trial will be convincing to that community. If it's not, then it doesn’t matter what the investigators think. So we always are shielding investigators to the extent that we can from the ongoing results.

And this is a technical complexity, but the overriding issue is always how convincing is it going to be to the community? And often, if -- literally everybody knew what the DSMB knew or what -- their equipoise would be disturbed in a trial. That's almost always the case.

But it's not the case that those same results at a distance will be so convincing to the community. So you can disturb the investigator equipoise and not disturb -- and still be in an ethically tenable situation. I just want to say that, which is not to say that this issue of not letting the adaptations be informative is not an issue. It is, but it's not that fragile.

And the last thing I'll say is very often the adaptation is done on safety. I'm sorry, on advocacy.

But the clinical decisions have -- and I agree with Roger's point that the physicians are not that sensitive to the size of the effect. But what we're all sensitive to, and they are actually sensitive whether they explicitly put it this way, is the effect in comparison to the safety.

And the safety often does not come into their randomization -- into the adaptation. So it might, it might not. So we always have to keep in mind that, in the end, the therapeutic decision is going to be based on some sense of what we're buying for the degree of safety risk. And the safety risk may or may not have anything to do with the adaptation.

If it is going to be a critical factor, then it should be brought in if they're on equivalent time scales. But the challenge, of course, is adaptation has to -- you have to adapt on the basis of accruing information and safety may be on a totally different timescale than the efficacy information, and it can go in both directions.

So it may not be possible to adapt on safety, and yet that might be the determining factor at the end

of the day. So we have to always keep in -- understand where we're getting the complexity from, which is very typically from the -- on the efficacy side, but know that the therapeutic balance is being judged with another parameter that may or may not be incorporated into the design.

DR. PRICE: Dr. Chan?

DR. CHAN: All right.

DR. PRICE: And please cut your mic off, Dr. Goodman?

DR. GOODMAN: Oh.

DR. PRICE: Thank you.

DR. CHAN: Okay. Thank you. Just wanted to add a couple points, and obviously I really liked the example of being sort of practical and not too complicated if it's really not needed.

For example, in many of the trials we conduct nowadays, we try to use just a simple futility rules as limiting -- or synthesize re-estimation. It's a very simple tool. And those typically could address maybe 70, 80 percent of the adaptation that we do to try to help us to be more efficient in running our clinical

trials. So that's number one.

Number two is based on the questions that was asked. I think a couple of the principles, one is that definitely we want to try to see if there's a sound statistical principle that's behind the adaptation rules, and that would be the best.

And if we get too complicated, the designs, a lot of times we may not have the theoretical results and then we have to rely on comprehensive simulations, look at different scenarios.

So having those sort of really workout is critical to ensure that people will be convinced about the adaptive designs that we put in place and review us in the scientific community and even ultimately when we try to report the results in the label and to the physicians, how you characterize the trials and how you adjust the potential bias due to the adaptation. Those are really critical elements that we need to spell out in the protocol.

And so, and I know that there's a second question that I'll touch on a little bit. It's the pre-specification is extremely important interface so

that we can have perhaps an independent committee to review the interim data and to help the sponsors or the institutions to make adaptation. So the rules of adaptation and what's the implication, those I think need to be very clearly laid out in the protocol. So I'd leave it at that, those two additional comments.

DR. PRICE: We have a list. This is great. So I'm going to say who we've seen. Dr. Lieberman, Dr. Harrell, Dr. Toerner, Dr. Chau, Dr. Zhong, Dr. Price, Dr. Emerson, Dr. Barry.

So we will try to get through as many as we can, realizing we want to get feedback on questions two and three as well. And I can -- I can say those again slower. So we're going to move to Dr. Lieberman now.

DR. LIEBERMAN: Okay. Okay. Thank you. Just a quick comment. So we talk about the need to pre-specify, but I think we have to think about it. Are there pre-specifications and the adaptation that would eventually un-blind the team to the interim results?

Because even if it's a different group doing the interim analysis, the implementations of the adaptations will be implemented by the team. And if

it's really pre-specified and very detailed, all of a sudden we say, oh, this is where we're going if we're doing this adaptation. So I think there has to be some thought about that. Thank you.

DR. PRICE: Dr. Harrell?

DR. HARRELL: Yeah, and I have two points. First point is whenever someone uses the phrase operating characteristics, I hope what's not in everyone's mind is just frequent, just operating characteristics.

Type one and type two errors are actually not even errors. It's not even the right term for them. They're probabilities of assertions. And so, what we need, at least as important as those, is the probability that the posterior probability will be definitive at some point, either definitive for futility or efficacy or harm. And we need to calculate the probability of inefficacy, which is a lot different from the idea type one and type two errors.

Second point is what I think is fundamental to any sort of clinical trial, but especially when you're using response adaptive clinical trials, is to have a

precise high-resolution response variable.

So I've seen investigator after investigator struggle to -- in a uncommon disease, to get 60 patients in a study, and then do a responder analysis which makes the sample size effectively 30 or 20. And so, to collect 60 and analyze the information as if it was 30 patients is really statistical malpractice in my view.

And so the choice of the response variable is all important. You have to be able to adapt on the basis of high information. And that usually calls for a continuous variable or something that's ordinal with lots and lots of categories and the measurements are reliable.

So a variable like bone mineral density is a high-resolution, high accuracy variable that you can learn from very quickly, just as an example.

So the net effect of not doing that is that people are learning about things in adaptive clinical trials where the signal is very strong in a binary response situation. So the signal has to be very strong. In many cases, it has to be super clinical.

It's more than a clinically interesting or clinically useful effect. So just remember whenever you're using a binary variable, the chance of missing a real clinically meaningful effect is really huge and the adaptation may not be very reliable.

DR. TOERNER: Yeah. Thanks for the discussion so far. As deputy director for safety in the Division of Anti-Infective Products, I think the most important principle is safety. And Dr. Goodman and others, thanks for introducing the topic of safety.

But I wanted to highlight what Dr. Ashby had talked about in the context of first in-human adaptive trial designs, and we're seeing this in our pre-IND consultation program where sponsors are increasingly interested in a seamless Phase I to Phase II development program within the context of one study.

And Dr. Emerson, in contrast to, you know, efficiency in having multiple trials, we're hearing the opposite, that it's more efficient to have one Phase I/Phase II trial design.

But when companies describe this to us, invariably the Phase II portion in patients with the

disease for which the drug is intended to treat, there's a question mark about the dose and duration of therapy because that's derived from the Phase I studies of multiple -- of single ascending dose and multiple ascending dose in healthy volunteers.

So we would view that second portion to fall within a category that we would characterize as insufficient evidence to support safety of patients enrolling in the trial. So we actually would consider that to be a clinical hold, where we just don't have information to support safety and efficacy.

And so, while we have no trouble at all embracing a seamless Phase I development program of single ascending dose to multiple ascending dose, where you have pre-specified stopping criteria based upon the observations of adverse events, that type of adaptive trial design is acceptable.

We do have a significant concern when moving into patients with the disease for which the drug is treating where we don't have enough information yet on the dose, duration of therapy and observations of safety in the Phase I portion of the drug development

program.

DR. PRICE: I'm going to take moderator's prerogative. Dr. Goodman, did you want to respond to Dr. Toerner?

DR. GOODMAN: Well, I'd be -- yes.

DR. PRICE: Okay.

DR. GOODMAN: This may be more of a question. Are you saying that the Phase I portion in this continuous, seamless elusion is going to be shorter than if they conducted two separate trials? Is that why you're not informed? Is that the concern?

DR. TOERNER: The concept that's being presented to us is a more streamlined and faster development reaching Phase III drug development.

So that's the concept that's being presented to us, that's it's more straightforward to have IRB approval, for example, for one trial, instead of having multiple trials in order to reach Phase III development.

But what we're saying is we don’t have information to support dose and duration of therapy for patients with the disease in a Phase II evaluation. So

that's the concern we have with an adaptive trial design, moving from Phase I in healthy volunteers into Phase II in patients.

DR. GOODMAN: Phase I in healthy volunteers, not Phase I in patients. That's the critical difference, right? That's why you're objecting to the transition?

DR. TOERNER: That's correct.

DR. GOODMAN: Oh, I see.

DR. TOERNRE: In our world, Phase I is conducted in healthy volunteers. We are gathering pharmacokinetic and safety information --

DR. GOODMAN: I see.

DR. TOERNER: -- on the drug itself before you move into Phase II in patients.

DR. GOODMAN: I call that phase zero. So we have a phase shift here.

DR. CHOW: It seems to me I guess we are already beyond that question. Number one, I think that I would like to take this opportunity to share with you some of my experience on the adaptive trial design. I'm not speaking for the FDA. I mean, this is just my

personal experience before I joined the FDA.

At Duke University, basically my experience that we pretty much followed the following steps for the clinical trial studies that have been utilizing the adaptive trial design.

The step one is that we will first ask the PI, the principal investigator, of the commission to come up with a so-called wish list. In other words I think that we would ask commission exactly what's on your mind. What do you want to do?

I mean, usually I think that basically my experience in communicating with the commission, they have some hidden agenda. They would like to answer all of the questions, and then with limited data, some more clinical trials, something like that.

So really, this is a way of forcing them to come up with some kind of a wish list, exactly, I mean, what's on their mind. And then after that, the step two, based on that wish list, then we can determine what kind of adaptations should be pre-specified. I mean, the list is very important because, I mean, based on this one, we cannot propose some kind of very

complicated adaptive trial design with some pre-specified adaptation, also as Greg mentioned in his talk.

And the step number three, and then after all this is all said and done, I think that we will develop the plan for assurance of the data quality because I think that analysis based on the clinical data collected from the complex adaptive trial design is already important.

After all this is done, I think that step number four, because I think usually with a lot of the adaptation, we may not have the statistical methodology fully developed in order to reflect the last adaptation. So I think that we would conduct the clinical trial simulation to evaluate the operating -- I mean, the characteristics, as Dr. Harrell mentioned. This is under the condition that if the statistical -- I mean, the methods of full data analysis are not fully established.

This is pretty much what's my experience for the clinical trial study utilizing the adaptive trial design. Along this line, I think that I would go back

to the question number one. Pretty much I think in my opinion there are three principles, I think, for the adaptive trial design.

Principle number is that I think, in my personal opinion, the principle investigator should be in the driver seat, not a statistician.

But I think in the past maybe 10 years, I think that my experience working with a commission, usually I think the statisticians are all in the driver's seat to tell the commission what can be done, what is things that can be done, what cannot be done or something like that.

So I think that the first principle I would like to offer is the PI should be in the driver's seat, not the biostatistician.

The second principle has to relate to the quality and the validity and the integrity of the adaptive trial design. The principle number three, I think that this is extremely important, in my opinion. We never misuse and abuse adaptive trial design. In the past many, many years, I've seen this. I think people at least misused and also abused adaptive trial

design. We have to avoid that from happening. Thank you.

DR. PRICE: So before moving onto two, we'll cover Dr. Zhong. You're next. Dr. Price, Dr. Emerson and then Dr. Barry, and we'll move on to question two and three, which we might have touched on already a bit.

DR. ZHONG: I really like idea that a Swiss knife, don’t make the adaptive too complicated and to be practical. And I also like Dr. Chow's comment on the scientific validity and on what you have in decision-making to guide the decision rule on adaptive design and that's very important.

But I would like to just comment on Dr. Chow about who's in the driver seat. I think that for adaptive design, it's not who's on the driver's seat, it's the complete collaboration. It should be statisticians, the scientists at the company, as well as the PI and the independent data committee.

So this is complete collaboration. It require education, right? And it's not just let the PI to join it. I'm a statistician by training. And I would like

to allow my counterparts -- allow to collaborate, understand what the adaptation is, what the purpose is.

So I think we're not in the driver's seat. But we all should be in the driver's seat. It's not just one person in driver's seat. It's we all should be in the driver's seat, right? So I think it's very critical.

Now, to get back to the questions and discussion, I think the decision rule is very important. We have to set up a decision rule. Some is critical. It used to be pre-specified, and some may not be able to pre-specify. We may discuss that later, but I'm not going to talk about it now. But please be acceptable to some unplanned adaptation that will not jeopardize integrity of study. I think it is important.

And the education part is also very important because -- I mean, even with one SI, right, sometimes it is a lot of emotional factor there. People outside the statistical community do not quite understand what the decision rule means.

And a lot of emotional factors there, and the

company could make an undesired decision when they emotionally look at the data and went to regulatory and said, okay, we have a fantastic drug, give us approval. Or they could, I mean, prematurely terminate the trial. So we really have to educate all the stakeholders in this process.

And I would like to add a couple comments on Dr. Mehta and Dr. Goodman's comments. Regarding the safety in adaptation, in my experience I believe we can build the safety criteria into adaptation. I just speak for myself. So that's number one.

And number two, in terms of whether or not -- I mean, it can relate to who drives what? I mean, the PIs -- like that's not -- don't make a comment. Don't interrupt when the PIs get comfortable. I think that it kind of depends. I mean, if we educate the PIs well -- I mean, reasonable investigators well, then we are on this common page. And even though they are very comfortable, which they know the patient n, but remember one thing. I mean, our drug development is to improve their care and health, the care of the patients, provide the care options, treatment options

to the patients as soon as possible. And therefore, I mean we have a common goal there. And if we educate the PIs well, even though they can enroll tons of patients in, but we are kind of obligated to the public, that we may want to think about adaptation to allow us the opportunity to bring the drug to the patients. I will stop here.

DR. CHOW: Dr. Zhong, I would like to clarify one thing. It's what I mean that I think we are in the driver's seat, I mean it's the project lead. Definitely this is teamwork. I mean, the commission, without the statisticians, they wouldn't be able to do it. So I mean, this is definitely teamwork. So the driver's seat, that means the project lead. Thank you.

DR. K. PRICE: Great. Thank you. As it turns out, I'll be restating or reemphasizing a few points that I think have already been mentioned.

But a couple of the things I think are really important to ensure that the appropriate and effective use of these designs and one of those is that there should be the holistic planning of the drug development -- of the development of the drug. So it should be the

exception that we arrive in the confirmatory setting and go, oh, oops, we need to do some adaptation here. These things should be thought about well in advance and part of the drug program.

And as part of that, thinking about what other studies might need to be accompanying it, if the adaptations that would be necessary would require some longitudinal modeling or we've talked about natural history studies or other things that might need to be happening in the meantime prior to the confirmatory portion of an adaptive design, those should be part of this holistic drug development paradigm so that we are most effective in that confirmatory setting.

The other point I wanted to emphasize is it is in fact a very cross-functional and cross-stakeholder activity. So those points have been mentioned. But it's very important we keep in mind that through the pilot program, and as we are proposing these designs and having conversations with FDA and industry and others, that there are cross-functional representatives as part of those discussions and that we are able to bring to bear patient perspectives and other

stakeholders' perspectives, and then thirdly, that as has been mentioned, I think to ensure these are most appropriate and effective is keeping the patient first, and what information do they need to make well-informed decisions about whether or not to take a specific therapy. Thank you.

DR. PRICE: Dr. Emerson?

DR. EMERSON: So just a couple of comments, just to focus on what we're talking about. Clinical trial development has always been a sequential adaptive process. What we're trying to do is speed it up now. We're trying not to go through wildly separating times of Phase II, Phase III.

However, that is largely the sponsor's view, not public health's view in the sense that we have lots of treatments being tried and if you study one and then there's the white space that's dreaded, you go on to study something else at the same time. And trying to optimize that is a very important thing.

So I understand Dr. Toerner's point. The sponsors all want to hurry it up, but I'm saying, you know, we don’t necessarily want to help them without

being certain that we're doing the same things that we could do, which is how many people review Phase II results as they then decide to put patients on Phase III currently? And are we trying to shift it that only three people know those results, rather than it's been published in the literature and people have thought about it and it's reviewing the thing?

There's no question that going through 70 different IRBs is a slow process, and don't want to unnecessarily go through that. But we need to make certain that we're keeping the population preeminent in my mind, not just the patients who are on the clinical trial, but how fast we can adapt things.

And then I will note -- and speaking to your question of we can't pre-specify everything. And the FDA is of course well-aware of that because they all the time reduce, restrict indications at the end, you know, that they -- without gathering more data on that, they suddenly say, well, we don't feel comfortable in the really light patients or the really old patients, or people with renal failure, and you have to do something.

And of course that I justify with saying it's the game theory, you know. The sponsor has more of an interest in getting this drug approved, whereas the FDA sometimes steps back and says we'd rather go with errors of omission rather than commission sometimes. And so, again, it's how can we -- vast improvements can be made. We just don't want to throw out everything.

DR. BERRY: So I struggle with -- and maybe it's why I live along this, that -- let's not be too complex. I mean that statement is a tautology. Too complex is too complex. The question is what's the right complexity.

Many of our adaptive designs, for example, Scott, are you need to slow down. We need more exposure. We don't know the answer. And it's getting it just right, what's the right sample size, what's the right time, what's the right complexity.

While you may do just futility, there are situations where much more complexity is the right answer for the particular situation. So I think it does go back to Dr. Ashby's point, that what are the right questions for the population, for this drug, for

the whole class? What is the right design that fits this?

In order to answer that, it's about simulation and pre-specification. The Swiss Army knife is nice. But I take it as a incredibly different point that I think it was presented. That's for somebody who hasn’t pre-specified what tools I actually need. I bring everything with because I don’t really know what I'm doing and I might need 200 different things.

If I know I need a screwdriver and I need a scissors and a toothpick because I've pre-specified and that's the right complexity for this, we're going to run better trials. We're going to get better answers.

And I think that's part of what we're here today, is in situations where the trial needs more complexity than we're used to, how do we do that well? And I think that's critical, is the pre-specification, simulation and understanding the questions and the threats to those questions that Roger brought up.

DR. PRICE: Thank you, Dr. Berry. That's a very nice segue into question number two, which deals with pre-specification. So we will ask you to formally

discuss the extent to which complex adaptive design should be pre-specified. And I see Dr. Mehta, followed by Dr. Lewis.

DR. MEHTA: I think that what's important to pre-specify in an adaptive trial is the method that you're going to use at the end of the trial for a final analysis. As an example, if you're going to drop a dose or increase the sample size, you specify in the statistical analysis plan that I'm going to use this type of method, control of error rate or combination or closed test. This is what I'm going to use at the time of the final analysis. That's all you need to specify.

At the time of the interim analysis, the actual decisions rule that you use to drop the dose or to increase the sample size, there can be flexibility in that. That will not affect the type one error, as long as you have specified the manner in which you will conduct the final analysis at the end.

I know that the regulatory agency wants you to also specify the decision rule that you will use. All I'm saying is that may not be the optimal thing to do because you might see things -- the DMC might see

things that make it a better idea to adapt in a different way. And as long as you keep the principle of how you're going to do the final analysis properly locked up, then I think that should be allowed.

DR. PRICE: Dr. Lewis?

DR. LEWIS: So I now regret putting myself right after you because I'm going to unfortunately contradict you. So in my view, the discussion that we are going to have about simulation is critically important because simulation often yields tremendous insights into the strengths and the liabilities of a proposed trial design.

And I find that people's intuition about the way a trial design may misbehave is actually not very good. And sometimes, in fact if you simulate traditional group sequential designs with some things like biased coin randomization, you can actually get operating characteristics, with apologies to Dr. Harrell, that are actually quite different than you were promised when you were taught these techniques early in your training.

So I think that simulation is a key thing we

need to be able to do well if we're going to be honest with ourselves about whether we under the behavior of the trials we're proposing.

To simulate, you have to pre-specify. So if you're one who's going to simulate, you have to be able to write code that simulates the trial. You can't simulate the trial without pre-specification. That pre-specification should enable simulation that includes realistic effects that may affect performance, such as time to information, missing data patterns and the like.

So in my view, for the context of this discussion which is complex, adaptive designs, my understanding, that simulation is necessary almost always. It's eye-opening when you do it on designs you thought you understood, but you were wrong, and to simulate, you must pre-specify. Thank you.

DR. PRICE: Dr. Bretz? And if I could remind everyone to please turn your mic off when not speaking. Thank you.

DR. BRETZ: Actually I think I can agree to both statements, in the sense that I would also like to

make the point that we should avoid pre-specifying specific algorithms and put them upfront in the study protocol.

We need to specify the type of adaptations, the methodology we're using to integrate the data and for the fine analysis. But that's probably about it, because at interim analysis we would like the decision-makers to be able to integrate the data they see and possibly move away from any pre-specified algorithms.

This does not exclude the need for running simulations and understanding the operating characteristics for some foreseeable outcomes at the interim analysis. So I would certainly like to advocate, if the teams come to me and ask for advice. I always tell them, so what is the interim decision you want to make and please think in advance what exactly.

So for example, in a treatment selection design, you tell me, yeah, we want to select the best dose. So what do you mean by the best dose? You have to tell that to me, right, because the DMC or whoever the decision-maker is will have the make the decisions.

And if it's an independent DMC, quite likely

they will have to make the decisions on our behalf, on the responder's behalf. So more importantly, we will have to understand the operating characteristics. At the same time -- and this will have to be communicated with DMC.

But at the same time, it allows the DMC to have the flexibility if unforeseen safety patterns come up, that they can read the balance, efficacy and safety and, you know, make the right decision, so to speak. So in that sense, I think I can agree to both of the previous panelists.

DR. PRICE: Dr. Meurer, followed by Dr. Ashby.

DR. MEURER: When I was first looking at this question, I thought that answer was quite simple, in short, which would be discuss the extent to which complex adaptive design should be pre-specified. And I would say fully.

There is nuance to this. But I think, you know, going to the last point, yes, data and safety -- an external data and safety monitoring board does have rights and obligations. But I guess the roadmap for the complex adaptive trial and the algorithms that are

pre-specified, I would sort of challenge, you know -- I think one of the things before was like we need to, you know, sort of respect the PIs. You know, I'm a PI. I'm a clinician. I'm kind of dumb. I mean, like I don’t know the -- you know, and I think this --

DR. BERRY: You're so hard on yourself.

DR. MEURER: But I think that if we have a dumb question, we need to be told we have a dumb question. And if we have a dumb way of framing things, and that, you know, we have a lot of other concerns that aren’t encapsulated in our primary endpoint, then maybe we need to have some other endpoints so that we can properly value those things in terms of whether they're important safety outcomes or not.

So and I think, you know, potentially if this was something like, you know, the consort statement, it would probably have similar things, like at least for a fixed design in terms of making sure all of those at least interim stopping rules are very pre-specified.

So I think pre-specification is important for the transparency and the, you know, reproducibility of the research. But that's not to say that data

monitoring committees shouldn’t have rights as well to perhaps do things that are pre-specified. But I think that's different from the performance of the design and the protocol as intended.

DR. ASHBY: Thank you. I mean, William just used the word that, to me, is the absolutely guiding principle here, which is transparency. That's what I'm most concerned about. And so, in that context, I would be very much in favor of trying to pre-specify the algorithm. I think that has benefits when one's working up the trial because it's only when you actually pre-specify it that you really have those conversations about what is the PI trying to do, what's the overriding goal, how are the statisticians interpreting that, if you've got patient input.

I've had some very good input from those in some of these discussions. It may well be when the trial is live that some other characteristics happen as well, and so that may change it. But it should be completely transparent what was pre-specified, what were the additional factors and why. And certainly -- and I've spent my time on UK advisory committees,

European advisory committees. What you want to know at that stage is what was pre-specified, what else happened and why? But it's got to be completely transparent. And if you get that right, I think the rest falls into place.

DR. PRICE: Dr. Lewis, did you have your hand up? And then we'll go to Dr. Goodman.

DR. LEWIS: So this is a discussion that's interesting that I was having yesterday at the Wellcome Trust. We were discussing E6 and some issues that have to do with protocol amendments.

And I think the concept of what is and is not a protocol amendment is helpful in clarifying the discussion. When I state that a trial design should be pre-specified, in my mind that is the trial as you intend to conduct it if nothing unexpected happens. And that ought to be pre-specified, and it ought -- you ought to be able to simulate it.

If something comes up and the DMC sees a safety signal that was unanticipated, there was external information, which I know was excluded from this discussion, and they, in their infinite wisdom --

I use that term guardedly. I'm on many DMCs -- chooses to drop a dose or change an allocation regimen, that's a protocol amendment. That's a change from what you pre-specified. And it needs to be reviewed through the appropriate mechanisms assure that that amendment, that un-pre-specified change is appropriate from scientific, ethical and regulatory points of view.

So I think that we need to be very careful about being clear when we talk about things that you might do partway through a trial to distinguish those things that are pre-specified and are not protocol amendments versus those things that are post hoc unanticipated changes that are maybe appropriate, but need to be clearly identified as such.

DR. PRICE: Dr. Goodman?

DR. GOODMAN: Thanks. That was -- that eliminated 90 percent of what I had to say, and said a lot of things I wouldn’t have thought of saying.

The only thing I will add is that it's also possible to explore through the initial simulations the effect of deviations from the written protocol and that's as important. So you can say, well, what

happens if we don’t adapt on, you know, the randomization or what happens if the surrogate is not informative about the final endpoint.

I mean, all of the sensitivity to the assumptions can be explored. And in fact, many of the operating characteristics, however we define them, frequentist or Bayesian, are pretty resistant to, you know, commonsense modifications as you go along, whether they be unintended or even in response to things that you should've predicted.

So I think it's very important to build into the pre-specification what happens if we're wrong and if we made different choices. And even you can do that to some extent ahead of time and find out that, you know, things won't change that much. And if they are exquisitely sensitive, you need to know that.

DR. PRICE: Dr. Chan, Dr. Meurer and then we will move to question three very quickly before opening up for public questions and answers.

DR. CHAN: Okay. Thank you. Just a couple things to add. I think -- and I fully support sort of in terms of, in the complex design, do we really need

to pre-specify the general algorithms as a guideline for the DSMB or independent review body to understand how the potential implications. And a lot of times when we work with our DMCs, we try to mark up different scenarios of what would've happened if the trial go this way or that way, according to how the adaptation rules were written.

And then when -- and the DMC will make their recommendation. But there are certain aspects that the sponsor would have some additional information outside from the protocol that could help the sponsors make a decision after taking the recommendation from the DMC.

So this is really critical, as much as possible, to let the DMC through the DMC charter or the strategic analysis plan specific to the adaptation rules.

So then -- and that goes with something that we haven’t really talk about, is the communication aspect under Greg, in his slides, talked about the importance of how you're going to maintain the integrity and communication and firewall, who actually have access to the blinded -- un-blinded data and who's

the DMC communicating to the sponsor's personnel. So those need to be sort of also pre-specified to really ensure acceptance of the adaptive trials in the scientific community.

DR. PRICE: Dr. Meurer?

DR. MEURER: One other thing that I forgot to mention, one thing that one has to be careful about -- and again, I am a fan of transparency and openness. But at times, you may need to not have your enrolling investigators in say like an un-blinded study with multiple arms, there may be some details of the algorithm that need to be held back from them so that they're not reading too much into they have three people in a row in one treatment.

Not that they should read too much into that, but that could be the transmission of information about the algorithm, in terms of like if it's response adaptive randomizations. That could lead to operational biases and that should all be pre-specified. But it may not need to be public. And we do have one example of a trial where we know that response adaptive randomization is occurring. But the

enrolling investigators don’t know the precise ratio that it can go to. But obviously the un-blinded statistician knows that and has codified that prior to the initiation in the study.

DR. PRICE: Thank you. So we were overly ambitious in our questions, which is a good thing, which means we've had a lot of discussion, and it's been very informative.

We're going to be a little bit adaptive. We're going to go to question three, but we're only going to take two responses -- whoever feels most -- oh, wow. This is not going to work. I saw Dr. Harrell and Dr. Emerson. So they’ll respond to question three.

We'll then move to public questions and answers. And we have a lot more discussion, so I'm sure you will have opportunities to speak as the day goes on.

So question three: Bias and treatment effect estimation is currently less well-studied that type one error probability control in the context of complex adaptive designs. How important is the evaluation of the properties appointed in interval estimates? And

should adjusted estimates be included in labeling and reporting of results? So Dr. Harrell first, followed by Dr. Emerson.

DR. HARRELL: I want to just steal a moment to celebrate two things. The first is I've been on a lot of panels without enough gender diversity, and this is not one of them.

And the second thing to celebrate is we have sitting, side-by-side each other, the incoming president of the Royal Statistical Society and the incoming president of the American Statistical Association. That's pretty darn cool.

So my comment about this question is that when you use a priority distribution for adapting or for stopping or for your final evidence for efficacy and you use that same priority distribution at getting the posterior mean, the posterior mean will give you the right calibration for early stopping then other adaptation.

So we know that if you stop early for extremely high efficacy, the mean would be biased. The posterior mean will pull that to exactly the right

thing.

So if you don’t use Bayesian methods, Scott Emerson has the best software for helping you derive the incredibly complex sampling distributions if you're doing a complex design or just doing group sequential designs. It's very complex and it's hard to deal with, unlike the Bayesian posterior mean.

DR. PRICE: Dr. Emerson?

DR. EMERSON: So I would not say what Frank just said about me. But the thing that I would say is I do think, thinking about the Bayesian properties is very, very important. But I don’t believe you ever get the right thing unless you take the population of priors. And so getting a population of priors is very, very important, and I think that's an important thing to look at.

I think it is very important to do this bias adjustment. But very few people do it. And I have some guesses as to why it's not done that much. One aspect is probably the regression to the mean of only reporting these estimates when the drug is approved swamps everything else.

And so, you know, there is this idea. But I will just note that if you look at the medical literature, there is some resistance of people to take any sequential results. There's some papers out saying if you stop early in a clinical trial, you get reversal of your results more often than not.

And I think that's wrong. I think that if you control for the P value that is stated, the sequential trials with properly adjusted P values do better than a fixed-sample study with just a P value, but just realize that there is that aspect.

And then the last comment I'd make is interval estimates are of course very, very important if you're going to go with the frequentist paradigm and you're interested in non-inferiority and things like that. And they should be proper intervals.

But in the -- probably including it in the labeling, I'm going to make the rash statement that this would make a material difference really only with hazard ratios because almost anything else you did, you'd give the raw data, and people would do their own simple analyses using fixed-sample results. So if you

just give me the proportion of, you know, people who had response, people are going to quote those response rates and not do the adjustment. And I'm not going to be in favor of withholding the raw data in this case. So there's going to be a problem.

DR. PRICE: We will open the floor up now for public questions and answers, or questions. I'm not sure about the answers. Could I ask you to please give your name and affiliation?

AUDIENCE Q&A

QUESTION: This is Qing Liu. I used to be at DB1, FDA a while. And now, I'm working for Amicus Therapeutics on rare disease.

Now I have three points to make. The first one I think goes with any other designs with which we follow the KISS principle, K-I-S-S -- keep it as simple as possible.

We need to understand actually what it is actually we're doing, you know, why do we need adaptive design? Can we find any other approaches to solve the problem? We have to focus on the ball, think about the problem we're solving and then find a solution to the

problem. Maybe adaptive design is one of the approach to solve the problem. So that's my number one.

Number two, actually this is a responding to many of the comments and discussions that we -- you know, any time we have a novel design, adaptive design, we need to really have a inferential foundation to the design.

By that, I mean, for example, estimation. The estimation problem has been solved for simple two stage adaptive design in 2002 in the JASA paper me, Mike Proschan and Gordon Pledger. And that procedure has been actually applied to many different other situations. Now, so that's regarding estimation.

Now, I read many papers by Scott Emerson, and he actually showed in one paper that, you know, there's really not a whole lot of difference from Bayesian and the frequentist. I mean, if you have a Bayesian design, you can actually, you know, do some transformation and convert to a frequentist design and vice-versa.

So that doesn’t really actually help anything. And I hate to say, actually, you know, Thomas behind me

from John Hopkins University, the wording for -- the problem, the fundamental problem in an empiricist approach. For example, if I have a sample size of three, I can do a T test, and I'm going to have alpha value of 0.025, right? One-sided.

What about actually I increase the sample size to 1 million? I can still do an alpha test, 0.025. So the problem is if you think about science, the larger the study -- assuming there's no other bias -- and then, the probability of making any error should be goes to zero.

So having said that, the right foundation for clinical trial and the interest in general also for adaptive design is really an evidential approach that can actually include frequentist, Bayesian as well as Richard Royall's law of likelihood as a special case. That's has been actually developed, and I've yet actually, you know, to make more public presentations.

Now, the last point I want to make is actually integrity. And this is based on a real-world evidence, meaning a real-world experience, is that, you know, we talk about integrity, openness, transparency in terms

of the process, I mean, set up the DSMB and the data and everything else.

So what about actually one of the biases is actually reporting bias. So remember, actually, adaptive design can be a two-edged sword. So you increase the sample size. You can drive the overall statistical significance. But also it gives you the ability to find out, ah-hah, there is a treatment by subgroup interaction where the subgroup is actually meaningfully pre-specified.

Now, here is a question, that when this study is reported, it's submitted to regulatory agency, I mean -- well, because this treatment by subgroup interaction is not actually pre-specified in the ISAP, in the agreed framework with the FDA, does the company has the responsibility to say actually in fact we find something else? And also, does the FDA should go further to understand you know, well, I mean, there's something else negative about this drug, even though the overall statistical significance is reached.

Now, the next question is that when the company publish the result in medical journal, should

company disclose the fact that there is a treatment by subgroup interaction. In fact, none of the group -- it doesn’t seem to be there is any substantial evidence for efficacy. That's it.

DR. PRICE: Thank you for the comments. I think in the essence of time, we'll move to the next person at the mic. Thank you.

QUESTION: Hi. Tom Louis, Johns Hopkins. A couple, just two brief points, one on the monitoring plan. I think we have to look at it relative to history. I mean, O'Brien-Fleming monitoring with nothing else. It's still advisory. I mean, it's not as through you're locked in.

I think of all these as an aid to navigation, if I can use a nautical analogy, and that we should study them and explore and be very detailed, as much as possible, knowing that it's always just a navigational aid.

And then, another point having to do with that, screen, as much as I'd like to reduce the prominence of P values, I'd really like to route the word bias. It has such emotional resonance. I'm all

in favor of good estimates. I like mean squared error small estimates.

Any of you who have ever fit a regression equation and have stopped short of saturating the model, heaven forbid, are fitting a biased model. So let's get away from that emotionally laden word and talk about estimates that are good either with mean squared error or something else that doesn’t just focus on that one piece. Thank you.

DR. PRICE: Thank you, Dr. Louis. And the speaker at the second mic?

QUESTION: Yeah. Hi. I'm Russell Ray (ph), from Ocuvia (ph). I have a question. These designs tend to be a little bit more fragile than the fixed designs. For instance, if you have a treatment by region interaction, that could affect your operating characteristics, even in simple cases of a sample size re-estimation.

Is there any thoughts of how we can account for that and how much we have to investigate those sort of issues that might cause the outbreak characteristics to not be what we want it to be?

DR. PRICE: Dr. Berry has a response.

DR. BERRY: Sure. So first of all, I think the whole idea of the adaptive design it's much less fragile than the fixed design. Running a sample size to a fixed number, getting to the end and saying, oh, shoot, I got the wrong sample size. I had these issues, is much more fragile than something that can recognize it.

So there are threats to the trial sample size, the conclusions and it might be these various nuisance aspects of it. And as you simulate the trial and you simulate these different effects, you can see that, oh, no, if I do this sample size re-estimation and I actually have a large treatment effect, my sample size grows enormously because it misestimates the nuisance parameter and you can see that as simulated.

So this incredibly thorough simulation and the threats to failure typically bring about adaptive things and better adaptive things to prevent exactly that for the whole purpose of it being much less fragile than just going to a number and getting to the end.

DR. PRICE: Thank you. And our final comment or question before heading to break?

QUESTION: Hi. My name is Smita Asare. I'm the executive director of the I-SPY trials operations. And a couple of answers that I can give you for some of your questions, one other thing is in adaptive trials, which is unusual or different from your regular trials, is that the data has to be available now because you are real-time assessing.

And I think a lot of people forget that data is hard to get in a clinical trial, and you always have to be thinking about your inputs and having a lot of control about those inputs.

In the I-SPY trial, we do a lot with MRIs, pathology. We have real-time data inputs with recruitment and one of the points I would like to say is recruitment has an effect also on your algorithm.

If your recruitment is slow and you don’t get to those certain endpoints, it can have a big impact on your randomization algorithm. So things to think about is data, data collection, quality, review. We have remote monitoring on all of those data elements, the

number of the inputs and of course time.

So your simulations are great. It's just that they also have to account for reality of the patients coming into the trial. Thank you.

DR. BEITZ: Can I make one comment?

DR. PRICE: A final comment before we head to break.

DR. BEITZ: What I find is that some companies come in and they don’t really have a stat analysis plan at the time their protocol is in. And for the sake of doing an adaptive design, I would argue that you really need to see the stat analysis plan to make sense of the whole picture. That's my little punchline.

MS. BENT: Okay. Thank you. So we're going to move to break now. We'll be back at 10:30. And for those of you who have not ordered a lunch, please order a lunch from the Sodexho kiosk if you're going to need one because they need to prepare for that. Thank you.

(Whereupon, the foregoing went off the record at 10:15 a.m., and went back on the record at 10:37 a.m.)

DR. PRICE: We will go ahead and get started

with session two. The format for session two will be the same as that of session one. However, the FDA panelists will be rotating out throughout the day. So I would ask my two colleagues to introduce themselves prior to Dr. LaVange beginning.

DR. SRIDHARA: Hi. I'm Raji Sridhara. I'm the division director of Division of Biometrics V. My division covers all of oncology hematology products. Thank you.

DR. BEITZ: My name is Julie Beitz. I'm a director of Office of Drug Evaluation III in FDA, CDER.

DR. PRICE: And our second presenter of the day really needs no introduction, but I'll give her a brief one.

Dr. LaVange is professor and associate chair of the Department of Biostatistics in the Gillings School of Public Health at the University of North Carolina at Chapel Hill.

Many of you are well aware that prior to 2018, Dr. LaVange was the director of the Office of Biostatistics in CDER, and as Dr. Harrell mentioned, she is our 2018 president of the American Statistical

Association.

So Dr. LaVange, without further ado?

SESSION II: GENERAL CONSIDERATIONS FOR OTHER INNOVATIVE DESIGNS INCLUDING EXTERNAL/HISTORICAL CONTROL SUBJECTS, BAYESIAN DESIGNS AND MASTER PROTOCOLS

PRESENTATION

DR. LAVANGE: Thank you very much. And I am so excited to be back on campus in spite of having to go through security -- worse than the airport. No, and I'm very excited that this meeting is taking place. How do I start the slides? We can get the slides going. There we go, okay.

This meeting is the culmination of a couple years of negotiation and work with our pharma and bio colleagues, and under the Prescription Drug User Fee Act, number VI.

There is a project -- the complex innovative designs project, was part of the PDUFA VI negotiations. It was a project that was proposed both by the FDA and by the pharmaceutical biotechnology colleagues on the other side of the table.

There was a great meeting of the minds. A

couple people here today, Gracie Lieberman, one, was part of that background negotiation. And it -- the PDUFA VI project also overlapped very nicely with the commitment under the 21st Century Cures Act, which was passed by Congress.

And the two together, both called for a public meeting to discuss complex innovative or novel trial designs. And this meeting is in fulfillment of that. In addition, they both called for guidance work, and I'll say something about that in a minute as well.

Partially to respond to Steve asking why we're dropping the word complex in, in front of adaptive, the project is actually complex innovative design, of which complex adaptive designs, but other designs as well are considered a part.

And for purposes of the PDUFA negotiation, we defined fairly simply that anything that required simulations basically was complex. Anything that didn’t have a simple analytical derivation for hypothesis testing or some other way of making inference. Now that's obviously a simplistic definition, but it, you know, helped further their

negotiations.

And in addition to the public meeting which we're having today, the guidance commitments, there's also a pilot program which you'll hear about this afternoon, which is a very exciting part of the program.

So as mentioned, the complex innovative designs, we broke into two buckets, one, adaptive designs that may be complex. Not all adaptive designs are, as was pointed out, because they're either adapting on multiple factors or they require simulations to determine their operating characteristics, which as Frank Harrell pointed out, could mean a number of things, not just type one error.

And there could be other reasons why the adaptive designs are complex, but certainly all adaptive designs are not complex.

The session we just had this morning, I consider that a warmup session, not very controversial. Everybody was very polite. So with this session, we'll get into the more controversial topics, and I'm expecting a lively discussion. I'll be disappointed if

we don’t have one.

I will just say one thing about the complex adaptive designs. The FDA CDER and CBER issued a guidance in 2010 on adaptive designs, in partial fulfillment of both the PDUFA VI and 21st Century Cures Act. The adaptive design guidance has been revised instead of being finalized in its current form. It was revised to make it a little bit more clear what was expected of sponsors when they had a complex design.

And this was done by putting the principles to guide any adaptive design into the guidance rather than talking about this design's okay. This design's not okay. And that guidance is in clearance. Greg Levin presented on it this morning. He's one of the primary authors of it, as is John Scott, who is in -- will be presenting this afternoon on simulations.

So this session is on other designs. They may or may not be adaptive, but they're complex for another reason. And there are lots of different types of complex designs that we consider under this rubric of complex innovative designs.

They might be complex because they're

leveraging external data sources for one reason or another. Could be a control group, could be a natural history study to inform your analytic model. They might be complex because they want to use another criteria, or they use another criteria for decision-making.

They might base interim adaptive decisions on probabilities of future success with another endpoint, for example. Or they might be complex because they involve collaborative efforts, novel ways of sharing data in something that might be called a master protocol.

So I'll just roughly touch on each of these three types to get the conversation started. But they -- this is not in any way meant to be an exhaustive list of complex designs. There's many other ways you can go into complex designs.

And also, just the purpose of this discussion is to advise the FDA, to let the FDA hear from sponsors, academics, other government regulators about the different ways to approach complex designs, not necessarily to endorse or not endorse any of the

examples that I'm about to use.

So why would you even need to innovate? Well, you might want to innovate because you've got a very restricted population size.

We had a meeting yesterday on rare diseases, another public meeting which was just a terrific meeting where people talked about ways that we can facilitate and accelerate drug development in even very, very rare diseases where patients are such a scare commodity, or a scare resource.

You might need to innovate because you want to try to improve the decision-making during a trial to get an answer sooner or to quit the trial sooner if things are not looking good, and do this in a complicated way by predicting probabilities of success, for example. You might want to innovate because you want to optimize product development by getting sponsors to put their data together.

But -- and this is consistent with what Roger Lewis said in the first session, that the real reason to innovate is to make sure you get the answer to the question you're seeking. It's pretty -- it's not

desirable ever to finish a trial and not have an answer.

And then why the need for FDA guidance? Well, the idea of adding to the guidance is to better understand if CDER and CBER, CBER being a sign-on to our 2010 guidance, whether CDER or CBER accept innovative designs, what innovative designs are accepted, how do they base that acceptance to better understand if you are running a complex design, how you submit that design, if it involves simulations, how you interact with the agency, what about the simulations needs to be submitted to the agency. That's the topic of session three.

And then finally, to try and ensure consistency of the advice that's given by FDA and the acceptance by FDA of complex trials across the therapeutic area, something that FDA is sometimes criticized for.

So I'll just take the three areas that I mentioned as examples, the first being rare diseases. And really this is more or less a recap of things that happened at the public meeting or discussed at the

public meeting yesterday.

But in rare diseases, when I was at the agency, even before I came to the agency, I worked in rare diseases. There's an interest in several things to accelerate drug development there.

One is the use of patient registries or natural history studies, which may exist in rare diseases simply because the disease is rare. It's often easier to build a more comprehensive patient registry. And those are rich data sources that can inform a trial design and possibly even contribute data to the trial when you can't recruit enough patients to have a fully powered, randomized trial that you'd like to have.

Also there might be interest in borrowing information from earlier trials of the same drug or trials of similar drugs in the same class or with similar mechanisms of action. This could be, for example, borrowing control data from an earlier phase trial in the same drug to supplement the recruitment that you're able to get in your later trial.

There may be information on disease prevention

from a natural history study that you could use to improve your analytical model. I'll allude to an example of that in a few slides.

There might -- as I've already mentioned, might be prior information from earlier trials, maybe not just the control arm, but maybe the treatment effect itself. Is there information you could borrow in a formal Bayesian way, for example, to help with the power in a trial if you're not able to recruit enough patients to give you adequate power.

And then finally are there ways to work with the heterogeneity that's so inherent in many rare diseases and look at endpoints that somehow maximize the power in the presence of this disease heterogeneity.

So, and I'll give an example where this was a recent approval of Brineura in a very rare form of Batten disease. This is the announcement of the approval. It was a very rare disease. The sponsor was able to run a single run trial. Much collaboration and back-and-forth with the FDA statistical review team and clinical reviewers about how to get the information

needed to go in the label.

And the label of the drug pictured here actually has comparison of the primary endpoint, and this is a time-to-event endpoint for probability of no decline in motor function. And it's compared to an external data source, a patient registry, which is fairly unusual for the agency to take this as evidence for approval, and then also describe the effectiveness weight in the label.

So this is an example to show that we do have -- okay, first of all, it's not we. I'm not here anymore. This is just an example to show that FDA does -- is accepting of innovation. This was an extremely rare disease that came through Julie Beitz's Office of Drug Evaluation.

All right. The second bucket, Bayesian applications. Why would the agency ever want to go Bayesian? Well, there have been a number of proposals made to the agency or discussions within the agency, things we're on record as accepting or at least considering. A lot of applications in the area of safety data, and this might seem natural because safety

data accumulate over the lifespan of the drug, and updating the information about risk is something that can be done fairly nicely with Bayesian methods.

One example I cite here is the interest in using existing control patients for large cardiovascular risk studies to just make more efficient use of the high risk patients that are required in those studies.

And another example of course in oncology, early phase dose finding might use Bayesian methods, Some continuous reassessment methods employ Bayesian methods, as well as Bayesian adaptive trials that might base an interim decision on an accelerated endpoint such as tumor shrinkage or progression-free survival, for example, or response, to basically see if that treatment effect can predict the treatment effect on the clinical endpoint, which is usually mortality. So an adaptation might take place based on a Bayesian calculation of the predicted probability of success on the other endpoint.

And then I've already mentioned rare diseases. There's many innovations in rare diseases that are

frequentist. There's also innovations that are Bayesian. The example I just gave you was a frequentist model-based analysis with external controls. There've been other proposals to incorporate prior information from early trials for rare diseases. There might be a use for the information about disease progression in the analytical model. I've already mentioned both of those.

A third possibility is using shrinkage estimators, some kind of empirical base estimator in a rare subset, maybe a rare genotype of a disease, where you would never be able to get enough patients in your trial to estimate an effect in that subtype with just those patients alone. So some type of borrowing is usually needed.

And then, a very good example that was talked about in the rare disease meeting yesterday, borrowing data from adult trials to help you reduce the number of children you need to study in your pediatric trials, something the FDA has talked about publicly and has a case study for that was presented last year in a pediatric meeting.

And then, I have an example of a Bayesian adaptive trial. This was a trial in septic shock. It was a Phase II/III trial that had planned enrollment to look at any of four doses, try to find the -- try to optimize the dose to take into the Phase III trial, then run the Phase III trial and the final analysis incorporated data, all of the data, on the drug that was under study.

And this was a fairly innovative design. It included adaptive randomization. It included adaptive interim decisions based on predicted probabilities of success, and it also had the option to end the study for futility as well.

And then the third bucket is the collaboration. So collaborative efforts that would be considered innovative, I group these under the rubric of master protocol, but there are many types. The general idea is that you study multiple diseases, multiple patient subgroups that are biomarker-defined and/or multiple therapies under one overarching protocol.

We've actually heard from the audience about

I-SPY 2, which was really a screening trial to come up with the best treatment in different genotypes of adjuvant breast cancer. The Lung-MAP trial, another example on the other end of the spectrum, this was setup to give confirmatory evidence of drugs running in parallel under the same protocol but firewalled against each other and not compared to each other.

The DIAN-TU trial in a rare genetic subtype of Alzheimer's disease, a very innovative, collaborative trial with multiple sponsors contributing their drugs, looking at this rare disease and use of shared control patients among the sort of sub-trials in that, as well as use of a natural history study to inform the analytic model.

And then, finally, the ADAPT master protocol for drug-resistant pathogens and the development of antibacterials, which is hoping to get sponsors to contribute drugs so that some efficient use of patients can be made.

Basically or briefly in master protocols, there's opportunities to innovate when you setup the trial infrastructure. There's also opportunities to

innovate in developing a common protocol and there's many advantages, including common screening platforms, that identify biomarker profiles and then move patients to the appropriate randomization arms.

The kinds of complex innovative designs that we can see in these protocols might involve adaptive randomization or adaptive enrichment. They may involve use of external and historical control data, possibly in conjunction with concurrent controls, so that both control data sources are combined for the primary comparisons, possibly sharing control group across protocols or across sub-trials within a master protocol if it makes sense for the biomarker categories being studied and then possibly model-based analysis methods, hierarchical-based models looking at subgroups, tumor types and so forth.

And this is an article that Dr. Woodcock, the CDER Center director and I, wrote on master protocols. Raji Sridhara, on our panel, has worked with many master protocols in oncology, and she and I both worked on the Lung-MAP master protocol.

FDA has come out fairly openly endorsing the

use of these collaborative effort, seeing them as an efficient way to use patient resources and also really possibly the only way to go as precision medicine gets more and more precise. It's just going to become harder and harder, we believe, to run standalone programs in some of these disease areas.

And then I haven't -- in addition to all the examples I just mentioned, FDA worked very closely with our colleagues in NIAID on an Ebola platform trial. It went off with only one drug, but it was setup to have multiple drugs if they were available.

There was a reason to go to master protocols, a concern in dealing with this epidemic that the drugs that were available would be in a very limited supply. We needed to look frequently at the results, basically looking after every pair of patients to see if we could stop.

And the epidemic nicely went away before the trial made it. Dionne Price, who's moderating today, was one of the key authors and worked with Michael Proschan and Lori Dodd at NIAID on the design. It is a Bayesian adaptive trial, as I mentioned, with frequent

looks at the data. It used a non-informative prior, so it's been called barely Bayesian.

And I will say interestingly, if you Google this trial, it's called Prevail II. The trial ended with a posterior -- it stopped early, I think 72 patients. There was a posterior probability that the drug ZMapp was superior to the control. So it was the ZMapp plus standard of care contribute -- being superior to the standard of care.

The posterior probability I think was about 91.2. But if you look at the way the trial was announced, in almost every public announcement the result was given as not statistically significant, did not reach statistical significance.

So this was a case where you pre-specified the analysis in terms of a posterior probability. But somehow the announcements always turned it back to relative statistical significance, which I thought was interesting. But anyway, it's an example of a complex trial.

So those are examples. There's many other ways to be complex. The topic for today is whether --

not whether, but given that FDA under PDUFA VI and the 21st Century Cures Act is accepting of innovating trial designs, what types of these trial designs does the panel think would facilitate the advancement of drug development, particularly in areas of unmet need. I cite two here, rare diseases, antimicrobial agents where we're desperately trying to encourage more drugs that are good to be developed.

And then second, what factors impact the perceived acceptability of innovative designs. There's a lot of "anecdotal evidence", quote, unquote, out there about what FDA will or won't accept. And it's of interest to us to hear how we can get a more consistent message out -- not we again, FDA.

And third, are there other outreach or research activities, areas for collaboration that might further advance the use and acceptance of these examples that I gave, but also other innovative designs, with the common goal in mind that the idea is to make drug development better.

So with that, we have two distinguished panelists who will react. The first is Roger Lewis,

who's already introduced himself, and following Roger is Frank Harrell.

DISCUSSION

DR. LEWIS: Thank you very much. I've decided to organize my initial comments according to each of the three questions, so the first question about what type of innovative trial designs which facilitate advancement.

In my mind, the key is the matching of the design to this scenario, again addressing what the true threats to success are in the development scenario or domain. I think we need to address these threats and focus again on getting the right answer to the question, either to inform regulatory decision-making or clinical adoption of the therapy being investigated.

So to drive this, I believe that we need two thing. We need a catalog of options that are relatively small Swiss Army knives that can be used in scenarios that commonly come up, threats, unknown rates of enrollment, endpoints, those sorts of things.

But we also need to have an explicit room for creativity when the relatively better understood of the

complex designs actually don’t address the real threats to success or the program. And there need to be some guardrails around that creativity, many of which were already discussed. But I think they include the need for rigorous pre-specification of the design, if we're talking about a confirmatory space.

And the simulations that are used to evaluate the performance of that pre-specified design need to be realistic. They need to consider things like missing data, time to information, potential bias, secular trends, differences by sites and then also sensitivity of the design to violations of various assumptions, very similar to what Dr. Goodman pointed out.

I think we need to be explicit in the fact that there are always tradeoffs that must be made in design decisions and allow for weaker performance in areas that are less important to drive the ultimate use of the data and therefore not allow, for example, concern over in some settings, inflammation of type one error rate or concerns over bias and estimation trump really important advantages that can be obtained through adaptation that are more important in this

specific setting.

So we need to be transparent about where we're making compromises in our designs, and what it is that is driving those compromises in terms of improved performance in other areas that are more important.

Secondly, believe it or not, this is the second point, I think we need to allow the explicit sharing of information among subgroups. The point was made about a rare disease subtype where you may never have enough information in that cell and you may want to borrow information, for example, using a Bayesian hierarchical model.

I think there are many more situations in which we often think we have enough information, whereas if you use a shrinkage-based estimate, which I think is actually a better estimate, you find out that that can qualitatively affect the decision you might be able to draw.

And I think we need to further develop experience, both from a design point of view and a regulatory point of view with the interpretation of estimates that are from hierarchical models.

And lastly for this question, I think the point that Dr. Harrell made about endpoints is really critically important. We need to show greater flexibility in identifying endpoints that are both informative and capture what's important to patients and their families.

I think there is a tremendous inertia based on approved regulatory endpoints that is really stifling and that we need to consider both endpoints that can be more rapidly informative and therefore help drive efficient adaptive design, but also frankly drive final decision-making. That was my noncontroversial comment.

For question two, and I may have been a little sleep-deprived when I wrote down my thoughts here, what factors impact the perceived acceptance of innovative designs?

Well, my first two were precedent and rumor. There, precedent I think is self-explanatory. The number of comments I hear about what the FDA definitely will not accept that has never been stated by anybody who actually hasn’t been employed or formerly employed by the FDA, is just phenomenal. And I think it's

interesting for us to consider what are the actual human factors or organizational pathology that drives the incessant repeating of these rumors.

First, my next line I wrote the reliance on data-free opinions of regulatory affairs specialists. It is amazing to me how a team that is working extremely hard to come up with a good design can be shut down by a regulatory affairs consultant who simply says, based on no information whatsoever, the FDA will never accept that, and then the team says, okay, we're done.

And if we could do something to corral those people or hold them to some evidentiary standard, I think that would be very useful. I will never be employed by anybody who works in that area.

I think a real issue is the variability of review groups within the agency and disincentives to be flexible within the agency. It is my opinion that -- or my impression, I should say, that the statisticians working at the front line within the agency are very limited in the time that they have available to understand complex designs.

And that leads to a risk, or a perceived risk, in the sense that it's easier to resist a complex design when you really haven't had the time to truly understand it. And I think we need to find some way to support regulatory professionals so they have a time to do their job well when their job is necessarily more complex than it might otherwise be.

I think another disincentive is a lack of skill and experience of statistical personnel in academia and industry and within the agency. I think there is a need for a general effort to increase the availability of software that allows one to simulate slightly more complex designs so people start to understand what affects the performance.

And I do think that within industry there is a disincentive to be innovative. And this was explained to me by a statistician who I will not name, who basically said I don't really think this drug is likely to be very good. So if I design an innovative design that might actually give it a better chance of success and it fails because the drug's a dud, my design will be blamed and my job is at risk. But if I do exactly

the last three failed trials that we've done and the drug fails, my job is secure. And so, I think we need to understand some of the organizational incentives that exist within industry.

For question three, are there additional outreach or research activities that might further advance the use of these designs, I just want to endorse the point that these collaborative efforts, often driven by patient advocacy groups to design platform trials to focus on finding the most effective treatment for a disease as opposed to exquisitely pinning down the treatment effect of each individual therapy, I think that's a key thing that we need to emphasize and support.

And then finally, as I mentioned earlier, I think that if we subjected our traditional designs to the same level of scrutiny and simulation that we do for complex designs, we would learn some things about the limitations or lack of pre-specification of traditional designs.

The one that jumps out at me is how often I see a trial that uses an O'Brien-Fleming stopping rule

and there's nothing in the SAP that talks about overrun or missing data or whether or not they've worried at all about what the DSMB is going to do if they're just close to the stopping boundary on either side, by the way, or not. And if we require the same level of pre-specification there, we would be setting a more equal playing field for traditional and complex innovative designs. Thank you.

DR. HARRELL: Well, those were phenomenal comments. My comments are not nearly as interesting and they're more focused on purely statistical things.

But I want to start with a hypothesis. The hypothesis is that Bayesian decision-making using skeptical priors will ultimately be shown to work better than using historical data. And the Ebola trial might be a case study in that, and I would like to see somebody test that hypothesis.

So if you're going to be borrowing information, there are a number of concerns. And I tend to be more skeptical about historical data than the average person is. And I'm especially skeptical when control data only are borrowed and just want to

make a note in passing that anytime you're borrowing historical data for a control, it has to be covariate-adjusted. To get the average for the control from the historical data is irrelevant. It has to be completely covariate-adjusted.

So if you're going to borrow information, there's various approaches to doing it, and I favor the mixture of priors approach, which is called dynamic borrowing. But I think that's a really bad name because you're setting the amount of borrowing upfront by the mixing proportion.

And this approach has a lot of advantages. It's only dynamic in the sense that when you're done with the study, you can get the posterior probability that the underlying treatment efficacy is coming from which source of the prior. So you can solve for that after the fact. But otherwise, you're just setting how much borrowing to do.

And there's advantages in the interpretation because the amount of borrowing, or the mixing proportion -- you're mixing say a skeptical prior that you would make a study stand on its own with versus

historical data where it might be an optimistic result from adults that you want to use partially for kids.

The mixing proportion is the probability of applicability of the adult data. So you have a very natural way to state that and interpret it. And you can actually elicit that, not only as just a single number, but you could get a whole distribution of applicability from a variety of experts and factor that in when you're calculating the posterior. So there are a lot of issues in terms of borrowing.

I also want to mention that if you're borrowing, there's no alternative to Bayes. We just don’t have the machinery in the frequentist world to handle borrowing, and so Bayes really becomes necessary.

I want to turn to another sort of complexity, which is actually simple, but most people think it's complex, which is fully sequential designs. That's something that hasn’t been mentioned yet. And we don’t ever see them done except in sequential response-adaptive randomization, like the classic ECMO trial. But I'm speaking of sequential non-adaptive trials.

So if you think about the last time you've worked on a sample sized calculation, if you've ever done a sample size calculation, you know it's not science. You know it's voodoo much more than science, and you know it's based on making up things.

And David Spiegelhalter has written about this very eloquently, that when you're doing a power calculation, if it uses unobservables, you're making a lot of problems for yourself and you're usually coming up with a very unrealistic power calculation.

So he and others show how to do Bayesian power calculations that don’t use any unobservables, and you don’t assume a point for the effect, but you assume a whole prior distribution for the efficacy that you don’t know.

But once you get into the fully sequential game, you can get past all of the voodoo and you can be like a physicist. So physicists say we're going to stop experimenting when we have the answer. And so, with the Bayesian world, the final answer that you get allows you to ignore all previous answers that you had along the way.

So all the previous posterior probabilities that calculated one person ago, three persons ago, four months ago, they're all forgettable, and then you get your final posterior distribution. So there's no way nor means to factor in the stopping rule or the schedule of looks in the data.

So this allows you to look after every patient or look every week or every day, doesn't matter. There's no type one error to inflate. There's nothing that gets messed up with the posterior probability.

So if -- to me, if sponsors could get into incremental budgeting, you could be doing incremental trials all the time and not assuming that you have some real notion of what the efficacy is that you don't want to miss. So I see tremendous bang for the buck in fully sequential studies.

The last comment I wanted to make is about the simulation. I don't think the simulations, if you're simulating type one error, can ever know all of the things that you need to know to accurately estimate the type one error. I think it's futile.

A recent blog by William Briggs discusses the

lady tasting tea experiment, the famous one by Fisher. And he questions that you can actually calculate a P value from the data only given what Fisher gives us about the way the data were presented to the lady and not knowing what the stopping rule was. You might not even be able to calculate a P value at all.

And so the idea that you can know the intentions of the investigators when you're writing the code, as Roger talks about, the code needs to have everything you know about. But you have to know the intentions of the investigator. And the chance of your knowing enough of that to write accurate code is very, very low. And so, to me that's a futile exercise, although you can get in the right ballpark. I will grant you that.

If on the other hand, you're simulation's giving you Bayesian power, such as the probability that the credible interval will be somewhere or the probability that the posterior probability will ever exceed 0.95 for efficacy or 0.8 for harm or 0.9 for similarity, those are the sort of things you can simulate really easily without hidden assumptions.

And that really changes the way we think about the problem because the probabilities that you're simulating are prediction mode probabilities. Based on what you know, we're trying to predict what we don't know, whereas in the frequentist world, you're assuming what you not only don't currently know, but you can never know, which is that the null hypothesis is true or the alternatives is true with a 20 percent effect size. The Bayesians really wouldn't have much to do with that way of thinking.

PANEL DISCUSSION

DR. PRICE: Thank you. So we will move to question one. But we'll be somewhat adaptive again. So if you would like to respond to question one or respond to Doctors Lewis or Harrell, please let us know.

And we have a suggestion to try a different method. So we will try this for this session and see how it goes. Instead of notifying Robyn and I, if you could just place your tent up, we will call on you, your tent card. So I see Dr. Emerson's tent card.

DR. EMERSON: So I appreciated the comments

from both the speakers. And the things that I would just like to really applaud is Frank talking about the true Bayesian measures rather than something that I don’t think has very good foundations, which is predicting -- using Bayesian statistics to predict the frequentist result.

You know, what I want to do is I want to get the Bayesian inference from that. And then, to me, the the strength of the Bayesian approach is -- and this I'm going to go out on a limb and state, is that whenever you're borrowing information from another data set, it better be Bayesian.

I mean it's very important to incorporate your uncertainty in all of this, and that does far better than some of the measures that will just use a point estimate and try to advertise -- average that in.

I will say that in the rare disease thing that I am very nervous about, which is people hiding behind, oh, this is a rare disease and we can't get enough patients.

And I'll just note examples in the not too distant past where one company presented a large

clinical trial in the exact same disease and couldn't get approval. And another company followed with a much small clinical trial later and was able to invoke it was a rare disease, you know, 12 versus hundreds of patients.

And so, there will always be the motivation to say we have to go to this data. But I think the comments that were made about needing to really pull in all of the data, not selective data, is very important to do that.

That the simulations that are done, which those simulations again are not really in the Bayesian probability space. They're really in a frequentist probability space. But they're useful. You need to know that as a very important aspect to do. And then, I think the simulations that you do are fairly comprehensive and are far more comprehensive than a lot of the frequentist approach.

Roger's comment about if you look at a fixed sample design that's often done by assuming proportional hazards, no time varying treatment effect and things like that, it's not very robust. You've got

to model something that's outside that range. And I -- from what I've seen of the simulations you do, you're far better at incorporating lots of different probability models than we tend to do with a standard frequentist design.

DR. PRICE: Dr. Goodman?

DR. GOODMAN: A few things. First, you know, I think it's definitely an advance that, you know, we get the acceptance of -- particularly in real rare diseases, of information from other sources that we might not otherwise have used in the past.

On the other hand, we have to keep our eye on the ball, which is we want to get to the truth. And the people at the FDA know better than anybody how enrollment in a clinical trial should be actually an approved therapy because it cures so many people.

And it's very, very possible in that one, in Batten disease, that if you actually could get kids -- if that was kids or adults, I actually don't know Batten disease -- enrolled in the trial, you would not see the same fate as you saw in the historical control group, and you know that. But if you can't get them,

you can't get them, and you don't want to never approve drugs.

But I think the kinds of activities that could done to buttress these sorts of approved designs is one of two things. First of all, to do maybe as exhaustive meta-research as you can, looking at the fate of control groups or registry prognosis versus the prognosis in control groups of clinical trials. So just see a priori how different that often is and how reliable this can be.

Of course, you always end up arguing it in each particular case. But it's very useful to have that on a systematic basis. But perhaps as valuable, and certainly this is gotten a lot of attention in the safety realm where there's been a lot of, you know, ink spilled on lifecycle approaches where you're looking at safety as it comes in over the lifecycle of the drug, even after approval.

In this case, you're actually not sure about efficacy, not just safety. That's what's more unusual. And if there's any way to continually monitor the landscape and look at the natural history -- the,

quote, unquote, "natural history" of the disease on or off the drug and see how that has evolved, it would inform you both about that drug, but also inform you about the future, about using these sorts of control groups because while it's great that the FDA's approving these sorts of designs in areas of special need, it doesn't necessarily -- you don't know actually.

They don't escape the laws that we've learned for other diseases. That is, these decisions could be wrong, which is not to say they didn't use the right approach. But they could be wrong, and they could be wrong at an unacceptable frequency.

So the more we can learn prospectively about -- with the experience about the accuracy of these judgments, the better. I had something else, but I actually forgot it right now. I'll come back to you if I remember it.

DR. PRICE: Dr. Chan?

DR. CHAN: I just have a couple comments to add too. There's lots of discussion already in terms of the types of innovative trial design that we really

need.

One is obviously we talked a lot about the rare disease area in terms of borrowing data from whether it's registry of other types of study. But I think the same mechanism would also apply when we are studying a new -- maybe studying a new patient population or a new compound where we might have Phase I data. And when we are going into Phase II, looking at the proof of concept, how can we incorporate the prior trial's data into the decision-making at all.

I know in a lot of companies are already trying to do those kind of incorporating the data into the decision-making. So the framework with the historical borrowing definitely is something that is really worth a try.

Actually, I'd like to see more maybe in the confirmatory arena, how the historical data borrowing can be done because that involves a lot about how do you weight the control data versus the concurrent data. And we talked a little bit about the master protocol example when you have several new treatments being introduced into the trial, but maybe over time, over a

length of a year or two, and how do you consider those kind of control data when you -- in the first period of the trial versus the last period of the trial. And some of those are a more complicated issue. I think those are the things we need to spend a bit more time.

The other part is the biomarker. I know here in the drug maybe really more specific. We talk about precision medicines, but using the biomarker guide in treatment trial designs, I think those would also involved a lot of adaptation potentially and could potentially change the estimate during the course of the trial. So that would require more the complex simulation to help us to understand what's the implication, difference scenario primarily.

So those are the two types I can think of that would be really critical to study a little bit more of the potential. Thank you.

DR. PRICE: We'll go back to Dr. Goodman, who remembered his third point.

DR. GOODMAN: Yeah.

DR. PRICE: And I'll just remind everyone when speaking, please lean into the mic so our audience can

hear us all. Thank you.

DR. GOODMAN: Yeah. So we all lean in. Everybody leans in. Okay.

So this is actually just to amplify a point made by Roger, which is the importance of communicating priorities that are appreciated by the leaders to all the different review groups.

I'll just give a little anecdote of my own experience. Now, this is a long time ago, and I recognize it's a complex organization. But there was a company that was developing a diagnostic device. They proposed that -- they were in oncology products and they were told their trial was unacceptable. And they were literally told to consult with me. I don't know why. So they did.

We redesigned it, and all the comments were right. They were very insightful. They were right. We redesigned the trial. It was shifted over to another group not within oncology, and not only -- and in the second group, we were told that the new design, which was exactly with the prior group had been -- told us to use, was unacceptable.

They used new ideas that they weren't ready for, like predictive value. The review groups had never used predictive value before, and they wanted to see it in terms of sensitivity and specificity. I'll just leave it there. And ultimately, the trial that FDA itself had internally recommended was turned down.

So these can occur among the leadership. But I think the organizational challenge is how you get -- and I think Roger said this, how you get this understood by review groups who are the affected decision-makers.

And we're talking about levels of sophistication here, one, two, maybe three levels above where they might have been operating, you know, above their comfort level. And I think this is no small task for a organization as complex as this. So that's a task for the agency.

You know, whatever's decided here, you know, that has to penetrate down, or if it's outside of their comfort level, there has to be procedures internally, and maybe you've developed these. I don't know, where it's kicked up to a level where it is within the

comfort zone. So they don't feel that, you know, certain kinds of designs are necessarily within their purview.

That's an organizational challenge for you. But I worry that at the top levels, there's tremendous amount of -- you know, can be a lot of sense, flexibility and wisdom. But this may or may not penetrate to the levels where the decisions are being made.

DR. PRICE: Dr. Lee?

DR. LEE: Yes, thank you. I have three points to make. First, regarding what type of innovative trial design will facilitate the advancement of drug development, Dr. LeVange already mentioned like platform design and the master protocol, et cetera. I think really we need to do more of this type of a design in the sense that there are basically two components in terms of what we can make of that advancement, okay?

And the adaptive design is all about learning and confirmation. So there's two steps. You know, we kind of iterate between the two, and for example like

platform design or the master protocol, this type of design can be very efficient in providing initial screening of the -- you know, the efficacy and toxicity of the agent, right, or the drug.

So if we do more of this kind of platform design and it can be a continual process rather than a one trial, one drug and one population at a time, and it can usually increase the efficiency.

But this, the learning platform, and with that information, then we can move into a confirmatory platform by designing more specific kind of a smaller trial to further test the finding.

So what I view is that we need to do better in terms of learning and confirmation, and learning and confirmation, and, you know, that if we can provide the framework of that, you know, from the FDA's perspective, it would really help to move the field forward.

The second comment is regarding again the inferential framework. And I would want to emphasize again that the traditional framework of the null hypothesis significant testing and the P value-based

approach is very limited.

You know, we assume the fixed effect of a certain treatment effect either under the null hypothesis or under the tested hypothesis, and try to control type one/type two error rate, et cetera.

You know, we need to move beyond that, and really thinking about the distribution of treatment effect and not just the efficacy, but toxicity, et cetera, okay?

So I think that again much has been said about, you know, the different approaches, like -- but I think the right inferential framework is still very important. And we need to really go beyond our comfort zone in terms of more fixed kind of a traditional approach.

So that leads to my third point is that also has been mentioned, like Roger mentioned about we need to have increase in availability of the software tools. So at MD Anderson, that's one of our goal, is not just to really derive new methods, but also provide tools, you know, to learn and to implement those.

So we have two kinds of tools under

development. One is the downloadable software. You know, you can download to your own PC, and then run the simulation.

But recently, that we developed application using the shining tool, which is all you require is the browser, okay? So we have two kinds of tool. You can just Google MD Anderson software, download our software online, and then, you know, you can kind of -- these are all freely available.

And I know that we have our colleagues in -- you know, in different areas they develop tool, you know, like in academia or in industry setting. And I think these are all good. You know, once we -- once we put the tool in place and then people can just use it, and learn from it.

I have met some, quote, unquote, "thought leaders", and all these are the expert in the field. They have very strong opinion but based on very little data or based on very little kind of knowledge.

You know, I really -- it's a humbling experience to me, okay? So then when we start to learning about this new method, that they design the

novel design, okay?

I would like to consider myself as a student of this rather than expert of this, okay? Again, you know, I think you all need to admit that whatever we do, whatever we know, you know, that there are probably more than what we know, what we understand.

And again, you know, if you had the tool available to people and then we can learn from it. And we can kind of debate about what assumption to use and what kind of result we'll get. I think we'll do better.

DR. PRICE: Thank you. Dr. Berry?

DR. BERRY: So I really like the comment of Peter that he's a student of this.

I think to answer this question, all of them, and innovations that aren't here, the really wonderful place about where we are is I think after perhaps 70 years of using agricultural experiments in medical trials, and we haven't innovated the clinical trial, but yet we have phenomenal drugs. We have phenomenal understanding of disease. We have personalized medicine, that we are now getting into wearables. The

complexity of what's coming in the next 25 years are going to meet designs we don't know about yet.

We may see master protocols the major way in which this is done. It may be incredibly complex to better answer the questions. And I think we are all students in this emerging clinical trial science.

So to limit these innovative trial designs, I don't think anybody here wants to do that. And every one of them that LISA put out I thought was an incredibly nice solution for the difficult thing that they were in.

And the hard part is I think exactly Dr. Goodman's point, is you are the arbiters of that in some level, and to be able to spread that within the FDA to arbitrate what is appropriate and is it appropriate to answer those questions is a really tough place. And the FDA is moving incredibly in that, and that will be one of the challenges to meet that.

So this innovative trial design, to put shackles on it in any way, or to, as Roger likes to say, to fight the battles of previous that we've seen failures, when the world is changing so much in terms

of the innovation, the science, the medical that we've got to innovate the clinical trial moving it forward.

DR. PRICE: Dr. Sridhara?

DR. SRIDHARA: So I think, you know, Scott brought this up very nicely, that we have to keep in mind that no one clinical trial design fits every type of hypothesis that we are going to test.

So I don't think we want to corner into any one particular design. But there are some aspects that I think we can keep, you know, think about when we're doing any of these.

One is to have a good knowledge of natural history of the disease. If we can work towards knowing that better, understanding the disease itself is good. The second thing is we have to have measurable outcomes in reasonable time. And what are these measures and what would make a useful inference and that would actually benefit the patients.

If we can keep that in mind, a measurable outcome that will directly influence how the patient can feel or function is very important.

Then, you know, if we are talking about

precision medicine, then another thing -- another key thing to keep in mind is diagnostic assays that are being used and are they standardized. Often we see one that's used in the lab of the investigator and then another one coming from a manufacturer who may be manufacturing this, et cetera.

So I think standardizing some of these or some of the antibodies that are used, et cetera, is very key thing in some of these as well. So there are other external things that we need to think about before we get into this.

I also appreciate that, you know, this complex versus adaptive designs because you can think of very complex designs in, for example, hematological malignancies where you have induction followed by consolidation and then maintenance. And you could be introducing new treatment in any of these three, and how do you put this all together.

So this kind of a dynamic treatment strategy, and how can we test these? And those are very important questions too. And I think we have not addressed any of that type of complex or adaptive

designs.

They are adaptive in the sense that those who get induction, not necessarily all of them go into consolidation. It depends on that they do to respond to the therapy or go to maintenance.

So it is by practice of medicine and taking care of the patients, you've got to do this. And I don't think we have touched on some of these complex type of clinical trial designs.

DR. PRICE: Dr. Bretz?

DR. BRETZ: Yes. When I see the term innovative trial designs, I wonder whether this is confined to the discussions we had so far or whether we can even broaden it.

And maybe it's not the time today to talk about digital data and real-world data. But I think there's a huge field that we as a community, as a society are moving towards into digital drug development age, and especially the use of data science, technologies, methodologies, whatever they are. But they go well beyond statistics.

And I think that will also raise some

important questions about trial designs and many innovative concepts. I'm not sure again whether we're going to discuss these today. But obviously these amount of technologies may bring, or so that's what they're claiming, bring the trial to the patients, which obviously have many consequences.

We have novel endpoints which can be assessed based on available designs or sensors, huge volumes of data. And the endpoints could be very different, maybe imaging being analyzed by algorithms rather than by human readers.

I think that all also has implications on trial design aspects, which go well beyond the discussions that we're having here. So at least I thought I'd mention this, that we don't forget about this type of innovative trial designs.

DR. PRICE: So just looking at the time, this is a very good discussion, so we're going to continue. But at any point, if you want to answer any of the other questions, feel free. So we are going to move to Dr. Zhong.

DR. ZHONG: I completely agree with Dr.

Sridhara's comment there. I think that's the kind of trial that we should target immediately and especially like Roger pointed out early on.

I mean, in the area of like not so well-understood trials, just, I mean, this something we have to target now immediately. It's kind of bring up the public learning and awareness of the FDA's openness to innovation and kind of correct the incorrect perception in the industry or the consultants there.

At the same time, I think that I agree with the last two comments, where we should -- from Scott and Frank.

We should keep our options open, right? I mean, the technology changes so quickly in this space. Maybe in like a wearable device, a digital or even big data. And even the statistical methodology also evolves. And we all are students to the new development of the technology and statistics.

So I completely agree with -- even though we have some structural needs that we need to start the problem with, but we also have to keep our mind open on different type of innovative scientific evolve.

All right. So now, I would like to talk about like the brief point that I made about rare disease. O mean, I'm on the bio and also pharma working groups, and sometimes, I mean, people kind of think that we all stick to like a rare disease.

But think about in sometimes, I mean, there's some disease that not rare, right? For example, Alzheimer's disease and cardiovascular disease. But there's some phenotype or subtype of the disease within those kind of common disease that I think -- I mean, as a concept, I think we have to think about this is a small population, as LISA correctly pointed out early on. It's a small population. So I think we should be aware of that.

And then, in terms of platform trial design, there's also like at the development in the industry and also in the technology as well. In this space, I mean, there a lot of collaboration across the industry within different companies. And one company could own multiple entities.

For example, for the treatment of Alzheimer's disease, and I mean one company could own different

entities or markers.

And then, we talk about the platform trial designs or master protocols. On the surface, it looks like maybe from one company, but in fact, I mean, that one company may be working on multiple different companies, on multiple different entities or markers there.

So please, one thing to think about is, again, the world kind of changes, not for sometimes things like one company but it could be multiple company collaboration within one company or multiple companies.

DR. PRICE: Dr. LeVange?

DR. LEVANGE: Yeah. I just wanted to respond. These are really fabulous comments. And just to make sure, in case -- I thought I made it clear, but if I didn't, so that the examples I gave were things FDA has already reacted to.

And they were not meant to be only the things that we're discussing here. They were just examples. The intent of the complex innovative design project under PDUFA VI and the 21st Century Cures is to look at other complex designs. What is the future?

So that might include, as Frank said, real-world evidence, not something we were particularly focusing on today. There's another PDUFA VI commitment. Aloka's involved in with that on real-world evidence. But it certainly comes in, the types of data sources you can come in, the wearable devices, the types of designs that Raji's talking about.

All of these are -- you know, this is really wide open. And you'll hear about this similar this afternoon. The purpose of the pilot is to get sponsors to innovate basically.

The FDA is tied with what they can talk about. We can't talk about designs, we -- the FDA cannot talk about designs they are reviewing. They're confidential.

So until the drug is approved, which can be years later, you're not going to know that FDA accepted that design, or FDA can talk about designs that are presented at an advisory committee, but that's pretty much it.

So the whole purpose of this pilot is to open the door for sponsors to bring in more innovative

designs well beyond, you know, the examples I used, which are now, you know, a few years old and are getting stale.

And then the last thing, just in case anybody is holding back and wants to talk about this, the simulation part isn't meant to be limited to Bayesians. There have been a couple comments. Yes, Bayes designs require simulations if you want to -- well, they always require simulations. And then, they may require simulations for frequentist characteristics if that's what people think is important. But some people don't. Looking at Frank, there.

But there are other complex adaptive designs that have nothing to do with Bayes that still require simulations. And there had been up until the time we were doing the PDUFA VI negotiation some concern that FDA was not accepting, at least in some therapeutic areas, an adaptive design that required a simulation because it's very easy to make a simulation space where there's a point that you exceed alpha 0.05. And so, it's very easy to ding simulation or that particular design.

So the simulation aspect, which we'll talk about after lunch, isn't tied to Bayesian. That's a complex adaptation maybe on two or three factors, and how does FDA view those, and how do sponsors get those in the door.

That's a possibility for a pilot that you'll talk about also this afternoon. Bayesian trials are a possibility. Other real-world evidence, all of these things are open. I just didn't want you to think that you only could react to the three little examples I gave, so --

DR. PRICE: Dr. Marchenko?

DR. MARCHENKO: Thank you. I don't want to repeat what everybody said already. I think a lot of good designs were mentioned.

But going kind of back to what Frank said, it's not just about real-world data, where we see big data. As sponsors, we collect a wealth of data, and not to use these data, I think it's not right.

So where I see probably advancement of drug development with regard to designs would be, yeah, Bayesian designs or hybrid designs which can

incorporate the totality of the data and give us a better decision based on the whole data because where I see right now, we look at the whole problem data only when we do the submission and the issue change.

Then I just wanted to note with regard to master protocols, I don't think the hesitation of sponsors are necessary in complexity of designs. I think the hesitation of sponsors in the lack of full understanding of the design and what is happening behind the scenes, specifically maybe how fast the data can be available.

What are the decision criteria which affect their specific drug, because we do need to remember that their drug, which they give to these master protocol, is a part of some kind of development program. So we do need to make sure the benefit, even for sponsors, is greater than the loss when they contribute to those master protocols.

But there are no questions. Master protocol is a great idea and platform for us to try to cure diseases.

And with regard to the next question, maybe

I'm just too blind, but I think that the lack of final or revised draft guidance on adaptive designs bring a lot of speculations about what FDA actually accept or doesn't accept and how less well-understood designs are understood by the public.

And with regard to the third one very quickly, I think we discussed again specific master protocols, designs which definitely helpful. But then, there are others opportunities for us to collaborate through different scientific working groups which are available under different societies, so we can put together.

DR. PRICE: So, thank you. I'm looking at the time and want to allow a couple minutes for the audience. So we'll take Dr. Lieberman and Dr. Beitz, and then, I'll do a time check to see if we need to move to the audience.

DR. LIEBERMAN: Okay. Thank you. So just -- I don't want to repeat, but just again sort of data source -- different data sources incorporating in clinical trials. So the real-world data, I think it's almost like -- it leads to the topic of can we start looking at endpoints and how endpoints in clinical

trials could actually -- or the real-world endpoints be reflected in what we collect in clinical trials?

The other one is actually the sharing of the clinical data for the control arms. And I know there's a lot of initiatives of that. So master protocols are not always available. So if there's a way of really thinking what are the right methodologies to really incorporate the control arms from the other studies that were just completed or very recently, so really there's focus on data sharing.

And as I just talk about the next points, I think, you know, some of the perceived just operational complexities of these trials might be a roadblock on both sides. And then with master protocols, or even the more complex decisions that have to happen, and you have a DNC and data coordinating center, it's like the industry. I'm not in charge of a lot of these complex decisions, so that could be another perceived.

DR. BEITZ: So I was intrigued by the comment that Dr. Lewis made calling for a catalog of options that can be used in scenarios that commonly come up while leaving room for creativity and flexibility.

And I was wondering if any of the panel members had some ideas about how we all could move from individual case examples to something more along the lines of a catalog of options.

DR. PRICE: Before we moved to the audience, does anyone have a response to Dr. Beitz? Dr. Lewis and Dr. Goodman, and then we'll go to the audience.

DR. LEWIS: So this is a very partial answer to the question. I think it's tied to the development of software. So software packages have some defined options, and so the space of options within software packages essentially creates a de facto catalog.

And if the software has the capability to simulate the performance of its options under a variety of assumptions and violations of assumptions, it creates a natural synergy where you both have defined reproducible options, and a mechanism for evaluating the performance of those options, so I think that's a partial answer.

DR. GOODMAN: Well, maybe I'll give another partial answer. I think so much of this depends -- and this is going to be repackaging the prior comment -- on

knowledge, or what you think you know about the reliability of your data sources and what you think you know about the natural history of disease.

And the natural history of disease is often quite different than the natural history of disease detected through all these different means, whether it's through sensors, whether it's through medical records, the prognosis, the course of that disease is often quite different than what we measure, either in clinical trials or in other settings, and it's constantly changing.

I remember the ECMO examples. It's a famous one I remember in the way back when, going to back-to-back conferences, one where it was discussed by statisticians, and literally the next week I went to a pediatric conference.

The statisticians were all discussing the nuances of the trial design, given a background mortality of 80 percent. The pediatricians weren't talking about that at all. They were talking about whether it was optimal therapy, and the fact that 80 percent wasn't close to what they were observing. They

were somewhere like 30 percent mortality rate. And that would change all the calculations.

So I think that the challenge of new data and new data sources is that we don't a hundred percent know what it is the disease that we are detecting anymore. Measuring blood pressure once a month, which might be a given risk factor with well-defined characteristics, is very different than measuring it, you know, 50 times a day. That's not the same risk factor.

So the kind of research that would make a lot of this, these innovative designs more possible is continuously either meta-research or the research into the validity of these measures and also the natural history of the diseases or risk factors as measured in these different ways because it's not the same.

The more secure we become in that knowledge, the more secure we can build the foundation of these designs, which often assume a lot more knowledge than we sometimes have with them -- than we need to have when we design the clinical trials with the concurrent control group.

And one thing on the sharing of data, which is somewhat related, is it's very different if you're sharing data were you already know what the result is in the control group versus prospective sharing.

So if we're using control groups from -- historical control groups or control groups from other trials where we already know what the answer is, then the decision to share by itself is going to be contingent often on whether the fate of that group was good or bad. And we can't assume that that's a stable property.

So sharing prospectively is profoundly different than sharing in groups that have already been observed. And you can be sure that that sharing decision will be partly based on what will make the therapy look good.

AUDIENCE Q&A

DR. PRICE: Let's move to the audience. We have time for maybe three or four comments or questions per person -- one per person, please. And if you want to begin in the back?

QUESTION: Thank you. My name is Cathy

Collet. I'm an ALS advocate. I can't tell you how happy I was to see these presentations, especially in this last session. I think they can be extremely helpful for an extremely challenging disease.

My question, it's a very practical one though. I see this wonderful embracing of innovation. About a month ago, we got ALS draft guidance from the FDA that really did not embrace innovation.

What are drug developers going to do? What are sponsors going to do when they have actually conflicting guidance? And my fear is that the guidance that embraces the innovation that could really help won't be taken up because there might be some that is actually dampening that.

DR. LAVANGE: So, right. So we are hoping that the pilot program will be embraced. This is really a case where I think the pilot program could be embraced. And we'll talk about that this afternoon.

But the idea with a complex innovative design pilot is that sponsors that have an innovative design will be able to get additional interaction with the FDA. And in exchange we, hope to be able to talk a

little bit more publicly about the design.

I'm looking at the lawyer who's going to be on the panel this afternoon to walk us through this. But that was in fact the intent of the pilot, is to be able to approach the FDA with a design that's not standard, that may be something they haven't seen yet and get reaction.

But the other -- you're coming from a patient advocacy group, I believe. And so, the other avenue is that patient advocacy groups have worked through the CPIM, the Critical Path Institute. Dr. Chakravarty could talk a little bit more about that, and I'm sure you're aware of it.

But some of these patient advocacy meetings are exactly what spawns the idea of the collaborative efforts and master protocol efforts. And one of the advantages of those efforts is that the group organizing the master protocol can oftentimes get regulatory interaction and discussion at a very broad level, even before you know exactly which molecules may be coming in the door.

You can get feedback about the collaboration

itself, the design, whatever innovations might be there. That's in fact one of the advantages of those coordinated efforts. So maybe either one of those might help, and I'm sure Dionne, Aloka and others could talk to you more at the break.

DR. PRICE: And I think Dr. Emerson wanted to respond as well.

DR. EMERSON: Yeah. Just a real quick comment. We talk a lot about innovation, and really what we want is we want proven -- innovations that are proven effective. And you know, we are experimenting a lot in clinical trials. And it's not true that we're any better at thinking up something than the medical community is.

A lot -- most treatments don't work, and so a lot of those things that we might think work may not. And so, I think we do need to pay attention to that. And we do need to start making certain before we go too far on some of these things that then won't pan out. We don't want to degrade the performance of our approved drugs.

DR. LAVANGE: Well, and that's a great -- I

meant to say this earlier. Thank you, Scott. That's a great lead-in.

So what's innovative today is not innovative tomorrow. What's innovative at the FDA may not be innovative in academia or industry. It's beauty's in the eye of the beholder.

Sometimes what all we need -- sometimes all is meant by innovation, when I think about it, are just things that you don't have a precedent for the FDA having accepted it, that that could be the definition of innovation in some eyes, to the regulatory affairs people that Roger referred to.

If the FDA hasn't been on record as accepting of this design, which in fact they may have, you just don't know it because the protocol's still running, then it's innovative because there -- and there's no guidance on it.

I mean, that's the problem, is what does innovative mean here, and we're not necessarily talking about the latest cutting-edge design. We could be talking about something that's not innovative at all in the eyes of academia, but may be considered

regulatorily innovative because there's no track record inside the agency.

And just to reassure Scott, nobody's talking about lowering standards here. We only want good drugs.

DR. EMERSON: But in medicine --

DR. LAVANGE: We only want good drugs.

DR. EMERSON: -- certainly there's been plenty of innovations that everybody believed that were proven wrong, and I don't think we're any better in statistics.

DR. SRIDHARA: So I just want to say that, you know, for the person who brought up this issue, we hear you. And we heard from the rest of the panel here as well that there are differences in the way we give advice between divisions, et cetera, and we will strive hard to make it as uniform as possible. So we hear you, and we will work on it. Thank you.

DR. PRICE: And there are mechanisms, as Dr. LeVange mentioned, to still bring in innovative design. So for time's sake, we'll go to the second mic now.

QUESTION: Good morning. My name's Mat Davis

from Teva Pharmaceuticals. I want to respond to the very first comment that talked about the barrier that we have to overcome as sponsors to get over the barrier of precedence and rumor that we discussed. And I think that legislation such as PDUFA VI and the 21st Century Cares Act, as well as situations like this help us to overcome that quite a bit on the sponsor side.

I think the second thing that we need to overcome is the barrier of the amount of time that these type of innovative statistical designs take. And I think we're overcoming that with the advent of new statistical methodology, new software available to us. So I think we're making a lot of progress.

The third barrier I'd like to bring up that I don't know has been discussed today is the barrier that we face sometimes in trying to make sure that once we do get one of these innovative clinical trial designs agreed to with the FDA, that that agreement will hold throughout the lifecycle of that specific design.

Some of these designs can take years to accomplish, and you can see that sometimes we'll have statistical reviewers change. We'll have clinical

reviewers change, and then some of the opinions on the appropriate -- and of that type of trial design may change by the time we embark on the trial to the time that the design end.

So I would be interested, if not comment now, but to think about later, what types of potentially new avenues we could have to ensure that once we agree on a trial design at the outset, by the time the trial finishes it will still be as appropriate as it was when we designed it in the first place. Thank you.

DR. PRICE: That is great feedback which we will take into our thought process. And we do strive for consistency. And science does evolve, but you're right. We do need to think about if you come in with a design over -- it does take time, and over time, the considerations that that design started prior to maybe changes in the science.

And hopefully with the pilot program again, I hate to sound like a broken record, but we will have discussions throughout the process. And I'll move to the third mic and then the second, and we will then break for lunch.

QUESTION: Hello. My name is Mei Chin (ph) and I'm officially from CDRH, but now I'm the detail in CBER. So I just wanted to echo what Raji said about the device. When we designed the clinical trial, we needed to consider the diagnostic device which classified patient marker status.

Because the device measuring biomarker are not all subject to measurement error because they are not particularly for size, or they may not be particularly accurate.

So I think the -- so considering the device early at the design stage is very important because different device may identify different set of what patient publishes, such as biomarker past-through.

So a device which have poor measurement performance could potentially lead to the loss of statistical power and also dilute the treatment effect factor.

And a major challenge that we haven't been facing is that sometimes in the clinical trial at the baseline, a lab developed a test which often is not perfectly precise or accurate is used to classify the

patient marker status. So I think we have been seeing this over and over again in a lot of statistic issues. So that's just my comment. Thank you.

DR. PRICE: Thank you. And our final question from the audience?

QUESTION: Actually, that was really spot on, because, I mean, breast cancer is where we work I-SPY. And as someone had mentioned subtypes in rare diseases, it's sort of becoming difficult to look at our trial as the one fit model.

Patients respond differently to drugs and we're noticing the biomarkers are going to be something that we really need to find a path through the FDA for regulatory approval of how do we utilize our data in a trial moving forward to get regulatory approval of those biomarkers in conjunction with those drugs.

And it doesn't seem like there's a very easy path to see, as far as like how do we do this. Two, we don’t just work with U.S. drug makers. We work with international folks, and how to get these innovative designs and the data from this accepted pretty much across the world.

How do we, you know, get a little more, you know, standardization, talking about the patient-reported outcomes? How do you get these pieces to work in parallel? How do we utilize this data, and frankly make it work both in the biomarkers, both in international? That's the questions we're dealing with.

DR. SRIDHARA: So we do have collaborations with foreign regulatory agencies. There are MOUs, and we do discuss with them on topics which are in some areas where it is not related to any product, but just the methodology itself as we do with the statisticians.

But within, for example, in oncology, we have regularly monthly meetings with six of the regulatory agencies where we do discuss with them specific products and what we are seeing and we do exchange our views.

It's not to say that we influence them in their decisions. We all have our own regulations and we have to follow them. But we certainly exchange our views and the way we are thinking about a particular method or some of the data that we are reviewing. So

it is happening, and we do discuss with regulatory agencies.

QUESTION: Thank you.

DR. PRICE: Thank you. This has been a great morning. We look forward to the afternoon. We will break now for lunch, and reconvene at 1:00.

(Whereupon, the foregoing went off the record at 12:08 p.m., and went back on the record at 1:09 p.m.)

DR. PRICE: Good afternoon. We will move forward into our first session of the afternoon. We have two new FDA colleagues joining us on the panel, and I will begin by asking them to introduce themselves.

DR. IRONY: Good afternoon. I'm Telba Irony. I'm deputy director in the Office of Biostatistics and Epidemiology at CBER, at the FDA.

DR. PERMUTT: Tom Permutt, Stanford drug evaluation and research. I'm associate director for statistical science and policy in the Office of Biostatistics.

DR. PRICE: And our first presentation of the

afternoon be given by Dr. John Scott, who is the acting director of the Division of Biostatistics, in CBER.

SESSION III: CLINICAL TRIAL SIMULATIONS FOR CONFIRMATORY TRIAL DESIGN AND PLANNING

PRESENTATION

DR. SCOTT: Thanks, Dionne. It's really a pleasure to be here today for this workshop that I think is really important and productive.

During the morning, there was a fair amount of incidental discussion about clinical trial simulations. In this session, we're really going to focus on that as a topic in and of itself, and try to hear some opinions from our panelists and from you the public about this topic.

So my goal is to give an overview of what we're talking about when we talk about clinical trial simulations in this setting and to raise some questions maybe for discussion.

So clinical trials have a variety of important operating characteristics and roughly what we mean by operating characteristics is expected behavior under certain clinical, operational or statistical

assumptions. Those operating characteristics guide trial design and interpretability. And one way of estimating trial operating characteristics is to simulate large numbers of clinical trials and to observe their outcomes.

So in terms of when you reduce simulation for this purpose or why, it is true that for many clinical trial designs, including some complex designs, statistical theory is available that provides estimates of important operating characteristics or at least bounds on those characteristics.

But you might prefer to do simulations or you might need to do simulations in at least a few different kinds of cases. One, if you're talking about complex designs that have multiple adaptations, statistical theory might not invite you with estimates that you can really use.

Bayesian trial designs, as we discussed in the morning, often, maybe always, require simulation. And we also might want to use simulations in small sample designs such as you would use for studying rare diseases, because the asymptotic theory that tells you

what happens as sample sizes get large can be unreliable with small sample sizes.

So this is a summary of some of the kinds of things we're thinking about or talking about when we talk about operating characteristics. So there's traditionally a lot of focus on the type one error probability of a hypothesis test. That's the probability of rejecting and null hypothesis that's actually true.

In addition to the type one error probability in the sense of mostly concluding the drug is effective when there's literally zero effect, you might also be interested in falsely concluding it's effective when there's an effect that's smaller than some minimally, clinically interesting effect.

Power is very important, expected sample size for studies that have variable sample size, group sequential, or other adaptive designs.

Estimation properties, we don’t always put in the category of operating characteristics. But because simulations can help in this area too, we sort of grouped it together. So this would be things like the

mean squared error of estimates, maybe the bias, if Tom will forgive me for saying bias.

And then, most of the things I mentioned above are primarily for frequentist statistical designs. If you're talking about a true Bayesian clinical trial, you might look at some alternative operating characteristics such as the Bayes average error or the maximum posterior probability of the null hypothesis in a rejection region.

So this is -- I focused on this slide on type one error probability. This is the logic of how these simulations work. The logic is pretty much the same no matter what you're trying to estimate by simulation.

So, but for type one error probability, you would start by assuming the null hypothesis is true. There might be many different ways of making that assumption. You would generate trial data under that hypothesis according to the design of the trial.

You would apply the trial analyses and the decision rules to that data, repeat that process a large number of times and then the proportion of times that that process led to a conclusion of effectiveness

would be an estimate of the type one error probability of the trial.

So these are several complications or areas where questions arise when you're talking about using simulations in this way, and I have a slide focusing on each of these separately afterward.

One is defining the null space, or more generally defining the simulation space. Two, is the scope of the simulations, how detailed, how many, how much. The third is -- I'm calling it multiple testing or multiple hypotheses. Really it could just be called complexity of decisions is a complication with simulation.

The fourth is applying these ideas in Bayesian settings with informative priors, and then finally, sort of resource issues and review issues.

So in terms of defining the null space, there are typically many different ways for a drug to be ineffective. We don't test what, you know, statisticians call simple-versus-simple hypotheses. There's a lot of different ways for a drug to not work.

So as one contrived example, suppose you're

studying a drug for a very aggressive cancer and historically you know that the median survival in this condition is about a year.

So when you're doing your simulation, the drug could be the same as the control. That would be the null hypothesis is true. There's no effectiveness.

But the control could have one-year median survival, as you would expect, or the control could have five-year median survival, which would be surprising in the disease, or mathematically it could have a thousand-year median survival, which cannot happen in human beings. But it's still in the sort of mathematical null space of the hypothesis.

So when we do these simulations, should we be simulating all possible null configurations or a sample of all possible null configurations or just the clinically plausible or important configurations? And how do we draw the line there?

In terms of the scope of simulations, in addition to an assumption about the treatment effect, you typically need to make assumptions about many other parameters when doing these trial simulations.

You make assumptions about clinical parameters. What is the true control rate in a trial? You will make assumptions about statistical nuisance parameters, such as the variance of an estimate. And you might make assumptions about operational parameters, such as the accrual rate to the trial or even the accrual per site in a multisite trial.

And so when you're talking about the scope, the total number of combinations of parameters is obviously infinite. So what kind of exploration of that space do you need to do to be reasonably assured that you have a good estimate from a simulation?

And then if you just look at one dimension, if you just look at, for instance, the control rate, you might want to explore control rates between 20 percent and 50 percent because that represents your uncertainty in a given trial. But should you do simulations at 20 percent, 30, 40, and 50? Should it be every five percentage points? All of these are sort of technical implementation questions that need a lot of work.

In terms of multiple testing or the complexity of decision, when we talk about simulations, typically

the focus is on the primary analysis of a primary efficacy endpoint.

But as we all know, actual decisions, actual interpretations of clinical trials are complicated. They depend on primary and secondary endpoints, possibly multiple primary endpoints, and they also depend on safety or risk.

And so, as we're talking about doing a simulation to get at what we think is our chance of making an erroneous conclusion, can we make simulations that encompass all of those multidimensional considerations.

I want to talk specifically about Bayesian settings. I also want to emphasize, as Lisa said in the previous session, when we talk about simulations, we're not only talking about simulations for Bayesian designs. There are other applications, other very important applications.

But the simulations become particularly important in Bayesian settings. And what we even mean by Bayesian setting is not always immediately clear. There are a lot of trial design proposals that use

Bayesian calculations, but rely on decision rules that are chosen to satisfy frequentist operating characteristics, to have a fixed type one error probability, and, you know at least a rough estimate of a type two error probability.

So when we're doing a trial of that kind, if it doesn't borrow prior information, the considerations are generally the same as in a non-Bayesian setting.

But when it does borrow prior information, the definition of the null space becomes really hard to get your hands on because your conditioning on data that have already been observed.

And if you're talking about the null hypothesis being true, you have to ask yourself were those data that were observed generated misleadingly under a null hypothesis or were they generated -- are they from an entirely different distribution from the new trial. And the way you set up those simulations is pretty unclear.

On the other hand, if you're doing a true Bayesian design where you're not looking directly at type one error probability, this would be a design that

follows the likelihood principal where inference is based on the interpretation of posterior probability distributions. These raise entirely different issues. It's still important to do simulation, but the interpretation is not the same.

So in terms of resource issues, doing simulations can be computationally intensive. They can be computationally intensive for manufacturers or applicants who want to include simulations in their poses and also for FDA doing review of those proposals.

On the computational front, there has been over the past 20 years dramatic and constant progress toward getting more sophisticated computational techniques and hardware. But there are still some problems that at least now, who knows in a decade, but at least now are essentially impossible to simulate.

So if you're doing a complicated Bayesian analysis with a lot of MCMC inference and you have to repeat that many thousands of times in a simulation, that can become intractable, depending on the specifics.

And as I mentioned, reviewing simulations can

be resource-intensive for FDA, and there are implications for FDA review timelines, workload, training reviewers who may not be accustomed to reviewing these types of proposals and also the software for doing the simulations.

One thing to consider, talking about simulations, is how do you report a simulation. How do you convey the findings of a simulation? And this is not an area where there's a template. It's an area where there's best practice sort of actively developing.

But some things that would be included in a simulation report would be a description of the trial design, some examples of hypothetical trial outcomes, if you ran through the trial a few times, the scenarios that are going to be simulated, the estimates of operating characteristics from each of those scenarios and then an overall summary of what this tells us about the design of the trial.

And in some cases, it may include simulation code, technical details for the simulation and statistical derivations.

So in terms of FDA's review of simulations, this is another area where we don’t have policy. We have developing practices which are not necessarily exactly the same in every review division or for every reviewer.

But some things that a review may include would be verification with the applicant's own simulation code or the off-the-shelf software that the applicant used for their simulations or it could include verification with code written by the statistical reviewer or with other off-the-shelf software to try to get a kind of second opinion. And it also may include exploration of additional scenarios that the applicant didn’t consider.

As of now, there's not a standard acceptance criterion for when operating characteristics are good enough or when our estimates are precise enough. This probably needs to be situation-dependent and certainly is now.

All right. So I'm going to run through the questions that we've setup for the panel to discuss, and then I'll turn it over to our lead discussants who

I think in order on the program were Scott Berry, Cyrus Mehta and Karen Price.

So question one, regarding the scope of type one error probability simulations, should all mathematically possible parameter values for which the drug is ineffective be included or only values that are in some sense clinically plausible? And if the latter, how do we define what's clinically plausible?

Question two, how should error rate simulations be conducted when formally borrowing prior information in a Bayesian framework? And what does type one error mean in that setting, and should we be considering other kinds of error rates more closely tied to a Bayesian approach?

And question three, what are some practical suggestions for implementing trial simulations, for example the number of simulation iterations, computational details or details about how this should be documented.

I also have -- there's a question four, which is -- I think it would be ambitious to think we'll get to it, and it's more technical, so we'll stick with

these three for now. So I'll turn it over to Scott.

DISCUSSION

DR. BERRY: Okay. Thank you very much, John. First I want to pause and reflect that how really exciting this is and what the question's FDA's asking, and the innovation that they're really pushing the industry for here of the clinical trial simulations, the adaptations, so just the questions being asked I think are -- is really very, very exciting.

Let me say something about simulations. The word simulations, I find that that term in itself is almost a dirty word, that we're used to this being predicting who's going to win the NCAA tournament, where is a hurricane going to go, these types of things, that it's forecasting and predicting hypothetical scenarios.

When we use simulation in this space, it's numerical integration. It's very simply numerical integration. We're not forecasting anything. We're calculating an integral and calculating various quantities from that integral.

The idea of being able to do those numeric

integrals using simulation as the general term for doing that, what it allows us to do is move designs away from boxes. We can do pencil and paper calculations of type one error and power in very restricted scenarios where the design is a box.

As soon as the design becomes any shape other than that, it's rare that we can calculate on pencil and paper those quantities. And hence, we do numeric integration. We calculate through simulation what are the characteristics of the funny oval-shaped design being proposed.

There's a beautiful aspect of this numerical integration/simulation that you can calculate amazing quantities that you can't do on pencil and paper in other situations. And I think it's had actually a strange negative consequence in drug development.

So for example, you can calculate in my Phase II trial, what's the probability this dose is selected as my minimally effective dose. You can't do that on pencil and paper. We don’t know that running a Phase II trial; hence, we run an amazing number of Phase II trials with three doses and 80 patients because we know

the power of a P value test of parallelized comparisons, despite the fact that nobody actually likes to do that in a Phase II trial. It's what we can calculate on pencil and paper.

Now we can simulate innovative dose response modeling. We can simulate the probability that a dose is selected and is moved to Phase III. What's the probability we make a go decision in a Phase II/III seamless trial? We can't do that on pencil and paper. We can calculate it exactly in numerical integration. We can calculate how much drug is going to be used in this trial, what's the distribution of that for drug supply.

So there's an amazing amount of things that we can then calculate. And then, we end up iterating the design invariably by simulating and investigating. And I was really glad John brought up the idea of looking at individual single simulated trials and showing that to the team and them seeing a trial run to the end and saying, oh, hey I didn't like that result. Oh, wow if that happened, that's a bad result.

It's amazing how often that happened, and then

that turns into an adaptive design. Let's mitigate that type of failure that we can see coming within the situation. And I'll give you a lighthearted scenario that happened to me more than once.

I take their -- the fixed trial given to me that has 80 percent power for a hypothesized delta. I simulate the trial and show them example trials, and that delta's a nice effect. And they look at me and say if I run that trial, 20 percent of the time I fail with my delta. Well, wait a minute. This is not a good trial design.

They didn’t understand what power was. Power to them is a threshold that means that the FDA's going to approve my trial. And a good statistician will get me a smaller n with the same power. It's a weird game, but simulation brings that out and shows it in complete detail of that.

So the idea of being able to use simulation, computer-aided design, is really a nice step forward and the review of course brings about various problems as John brought up, the various questions.

Many of these questions revolve around the

idea of type one error and the idea that has been engrained in our heads of strong control of type one error.

If there's an infinite number of null scenarios, our design can't be approved by a control of simulation -- by simulation because we can't simulate an infinite number of scenarios. And that sort of cycle is what many of this goes to.

So the idea of being able to simulate from the null, we do simulate from the null a great number of scenarios. There are strange scenarios where it's not clear if it's a null or not.

When you do enrichment designs, personalized medicine-type designs, and you could make a conclusion within different subsets, you might miss and hit at the same time. Is it a type one error and a type two error? Is it not an error? These are hard scenarios. They're really therapeutic to go through and recognize that these can happen, and what do we want the trial to do in these different scenarios?

So the idea of being able to simulate type one error and demonstrate that I think is wonderful. We've

presented many grid search scenarios. We take expected scenarios, and then we double them and half them. And we do four or five within there. And then we take dropout rates and we vary that over a scheme and show the characteristics of the trial with different dropout rates.

Sometimes enrollment rate changes the characteristics of the design, especially in an adaptive design. So we change enrollment rates within the trial and we demonstrate that.

An interesting thing happens with simulating a null scenario, is that we simulate -- and suppose you got it exactly right, and your trial is calibrated to a 5 percent error rate. And I simulate a hundred nulls. Half of them have elevated simulated type one error. They'll be above 5 percent. You'll be at 5.03 percent, 5.8 percent.

We know through the natural error of that simulation that we get above, so it's understanding that. So we might even have a particular delta that above this value is above simulation if we simulate above 100,000 scenarios from every null.

And we may do 300 null hypotheses with 100,000 simulations of each trial to characterize the probability of an incorrect conclusion in order to justify that the design is a confirmatory, adequate and well-controlled trial.

And I think as our simulations get more advanced, they get faster, our ability to do this is better, we can do much more of this within the adaptive design report and submitting these to regulatory agencies.

So in terms of the type one error probability scenarios, we can't simulate every possible value. I appreciate the question and it brings up this idea of theoretically understanding on every scenario. But we can do a very nice grid search for what that type one error is.

Sometimes we even do modeling of the results of the simulations to look at the effect of a factor, and it looks like that factor doesn’t affect type one error. And a model can demonstrate that on the results of the simulation in order to give further understanding of the behavior of the design.

I believe that there is no such thing as a design that's not well-understood if it's been simulated. You understand everything about the design as long as it's been simulated under that scenario. So we can lose that nomenclature eventually as we simulate trials.

One other thing I'll bring out about simulation is that we've proposed post-trial simulation. So we do a grid search which includes nuisance parameters, event rates, median survival. And we've done median survivals of six months, nine months, a year, 18 months and two years.

And then we can go back when the trial's over and bootstrap the trial over again, taking every patient from every arm equally likely, and it's a guaranteed null hypothesis. And what's the behavior of that, just for further understanding after the fact when the trial's over.

It's not something that, you know, you might even get a 2.6 percent type one error, and we can understand what that is. But it's further understanding after the fact of the trial, based on the

prospectively defined trial that was run during the course of it. All of this revolves around completely prospective design, full details done.

You can't simulate a trial that you don’t have the details for. And it's a really therapeutic way to create the protocol itself, is to ask a statistician to simulate it or a computer scientist to simulate it for you.

If they don’t know how to do it, you don’t know the design yet. So it's a really important part, and the various questions -- I agree with John completely, Dr. Scott, that it's a case-by-case scenario.

Some scenarios we've done 300 nulls. Some scenarios we've done eight, because of the particular situation, and that the trial aspect of it in there. Now, that makes it harder for the agency for sure.

It would be really nice if the agency would create a pilot program and investigate several simulated trials over the next year or so. It's a beautiful idea. So I look forward to that, and the exploration of the role of simulations in this. Karen?

DR. K. PRICE: I think it's Cyrus.

DR. BERRY: Cyrus?

DR. SCOTT: Dr. Mehta has slides, which I think --

DR. MEHTA: How do you move forward? I just have a couple of slides on simulations in frequentist trials and a couple of trials on simulation in Bayesian trials and then one slide on the computational issue that John brought up.

So in the frequentist setting, I think solutions are valuable for verifying asymptotic results. Asymptotic results exist, and for example in time-to-event trials, you will have -- you know the asymptotic properties of the log-rank statistic. But in small samples, it might not hold. You know very well what happens in nominal and binomial trials when there are a nuisance parameters. So it's very -- it's valuable to verify.

But if no analytical results are available, then it's essential that you simulate. For example, power, and then the type one error of a response-adaptive trial. You wouldn’t be able to get that

analytically and would have to simulate. And similarly for more complex designs by the arm trials with some type of adaptation and also multiple endpoints, if you want to know power, you may need to get that through the simulation.

Now, one more type one error control by simulation, it's not straightforward. Adaptive designs, multi-arm designs, multiple endpoints, as was just mentioned by Scott, they are complex null spaces that cannot be exhaustively explored.

And so you -- but fortunately, at least in the frequentist domain, we have very good tools. These tools have been developed over time, closed testing, combination of P values, preserving conditional error rates and independent increments.

These are tools that have been developed in the frequentist domain which do control the type one error in the strong sense, and so they -- I'm not really a case for no simulation. But you don’t need to worry so much about this more complex null space because it's taken care of by these tools.

Now, Bayesian is a much different issue. I

have not as much experience with Bayesian methods as Scott does, but I have some experience.

So I think a simple case that you just have a normal delta, mean and standard deviation of one. And the goal is to determine if this delta is greater than deltas -- than some null delta naught. Then the Bayesian inference retrieved this delta itself as a random variable and it puts sort of a criterion for success, which is this probability -- this posterior probability that the delta exceeds delta naught, given the trial data and given some historical data, should be good, should be greater than some gamma.

And then you can simulate -- you have to simulate that, that probability of the success criterion and under various assumptions about delta, how often will this success criterion actually hold.

And in particular, you want to simulate it under the null space which is in this case delta naught. How often will this success criterion hold if you simulate under the null space, and you'd like that to be bounded by less than alpha.

No, so DSIM is the data from the simulation

and D0 is the data from the historical trial. Now if you use a flat trial, then the Bayesian and frequentist results should be similar.

The issue is what happens if you want to borrow data and it's not from a flat trial. Then you have sources of bias. And there are two sources of bias. One is a selection bias, and that is like publication bias.

It's something that Professor Goodman also hinted at in his morning session, that well, you know, I've got all this prior data. And if this prior study was positive, I'm going to bring it in as my prior. And if in fact I didn’t have a prior study that was positive, then I won't bring it in.

Now, so that can be a source of bias. So there was a notion of only bringing in -- borrowing prospectively. That is to say, the trial -- the previous trial is still ongoing, and you don’t know its results. And then, you may or may not have this problem of a selection bias.

There are ways of contamination, which is that the control of type one error, it may be difficult

because the control arm of the prior study is worse than the control arm of the new study, and so that gives a break to the new study, or the treatment arm of the prior study is positive, but the treatment arm of the current study may not be, and both of these sources can inflate the type one error.

Now, so there have been ways of trying to avoid these biases. And one of them is this power prior method for controlling the alpha. This was proposed by Abraham and Chen, and what it does is it handles the heterogeneity between the current and historical data through a power -- through raising the likelihood of the prior to some power, A0.

And so this A0, so you see the posterior is part of the -- or a proportion to the likelihood of the new study and the likelihood of the old study raised to a power. And the extent of this power A0 determines the amount of borrowing that you can do.

If this A0 is zero, you cannot borrow at all. If A0 is one, you can borrow everything, and you can choose in between.

And then the question is how do you choose

this A0. There is an approach suggested by Hadad. I actually won't have time to go into it, but another option is to just pick a fixed A0 and then use that for controlling the type one error. And it sort of estimates the heterogeneity between the historical data and the trial data.

We did simulate this method of Hadad to see what would happen on the different levels of contamination. So you see these two tables. One is where you have heavy discounting and the other where you have mild discounting.

And each row is representing more and more contamination in the sense that the group, the new trial has a zero delta, no more treatment effect. And the previous trial has treatment effects. So no effect, effect with a mean of two and effect with a mean of four.

And we did simulations to see that under all these different situations, the type one error can be controlled -- or slight, almost controlled. Actually the alpha here should be 0.025, but the -- you can get close to it unless there's heavy contamination, in

which case, you know, you might get type one error of an almost 10 percent.

My last slide is about computational issues. If you're going to use these methods, you have to do a lot of simulation. And again -- and at least in the Bayesian setting, that might be computationally intractable.

This is a nice table that was published by Martin Posh and colleagues in 2011. It says suppose that the true type one error -- the true type one error of the test was 0.026. Then how many separate simulation runs would it take on average to get a simulated type one error of less than 0.025.

If you only have 10,000 -- if you have simulation runs of 10,000 then one -- on average, one in every four will incorrectly bring you below 0.025, when the true at this point is 0.026.

Is you have a simulation runs of 100,000, then about one in 43 times will it happen that falsely your simulation goes below 0.025 when the truth is 0.026.

And of course if you say I can simulate a million times, then it would be all right and you'll

have about 8x109 before you get a wrong answer. Thank you.

DR. K. PRICE: Great, thank you so much. Thank you, John, for a great introduction to this topic and for bringing out many of the important potential hurdles, as well as the challenges that we have, but with an eye toward action and toward identifying solutions. Appreciate that. Thank you for the other discussants, as well.

And what I want to do is take a little bit of a different approach. I think not so much answer the questions, but maybe just talk a little bit about some of the things that are on at least my mind, and folks in industry's mind on how we may be able to move forward together and that we're excited to move forward together on this.

First of all, as I mentioned this morning, I think it's important that we remind ourselves that clinical trial simulation to some may seem as a very heavily statistical activity, PK/PD activity, and it certainly is. Statistics features very heavily. But it is strongly a cross-functional endeavor.

And what we have seen many times, and has been mentioned I know by Roger and Scott and others on the panel, is the tremendous amount of unexpected benefits that we are able to come to through that cross-functional dialog that ensues when we simulate trials.

And it has been argued, and I had planned to also indicate that really any trial can and should be simulated. At Lilly, we do simulate, or at least endeavor to simulate every trial that goes out, regardless of whether or not it has a close form solution, in large part too, look at those unexpected things.

As was mentioned, intuition doesn't always play out. And so, we're able to see areas where the design may break down, and greatly improve that design. So just something I want us to keep in mind as we go forward, is the cross functional nature, and I'll come back to that before I conclude.

Also many of the designs, obviously as we're talking about today, require simulation. And so, what I wanted to do is think about it in three main points.

And the first thing is that we can do this.

We must do this, and we're ready to do this. And so, when I talk about we can do this, what I mean is we have tools. We have knowledge. We have examples. So we can do this.

Sometimes we probably get in our own way, and get a little caught into the details and concerned about things that we can overcome. We have, as I mentioned, tools. We have knowledge. We don’t have a tremendous amount of experience, and that's what this pilot program is so well-suited for. But we should not let that get in our way.

And we certainly need to improve on the tools. But I think that it's through the experiences in the actual application and just moving forward and just doing this that we're going to figure out where do we really need innovation in terms of the tools. Where do we need innovation in terms of the methods and so forth?

So we can do this. What we really have needed though is a pathway for communication. So what has happened historically, and we don’t want to spend a lot of time talking about the past, but learning from the

past is that we might conduct a lengthy simulation study. And by the time we have an opportunity to have a conversation, it's well into time for the study to be conducted and maybe too late to address concerns from FDA's perspective.

And so, having avenues and clarity on those avenues for communication will be vitally important. And I know that is a component of the pilot pathway, or the pilot project.

But we really do want to make sure that we have clear clarity on that, as well as a timely communication and understanding of if there are rejections for the pilot, which just may be because of other similar things have been proposed, there's clarity on that to the company so that we don’t revert to interpret that to be some negative view of the study and have a pathway to continue that development.

So we know that innovative designs are going to need to continue even outside of the pilot program.

We must do this. I think that patients demand it. We're all patients at one time or another, and as we talked about, to get to the truly innovative design

and the best thing for patients, we're going to have to do this.

And along this line, I think this is where as I was getting at earlier, sometimes we get in our own way on this and want to be careful that in order to do this is bringing the cross-functional groups together.

As statisticians, we can get into the technical details. There was a conversation earlier about regulatory scientists who certainly can be slowing things down in some sense. But I know some really great ones, and they can be some of the biggest advocates, and could really be essential to helping us more this forward.

And so, I think that bringing everyone together in these conversations will be important. Ensuring that we have medical reviewers onboard and that are part of these conversations so that they're onboard, and there's not a perception that this is only one group that is supportive of these approaches will be key.

And finally then, we're ready to do this. I just wanted to reiterate that we're excited from an

industry perspective to collaborate together. We want to look to develop best practices together. I think there are attempts at doing this. Let's look for ways and understand better how specifically can we do that. Can we leverage the scientific working groups?

Can we do other -- have other meetings, or other conversations to develop those best practices, the timely communications, the other ways that we can improve these, the ability for us, help us understand what it is that you need.

Help us to understand what we can give you, and how we can do that in a standard manner so that you're able to more efficiently and effectively review simulation results and ultimately approve these designs. Thank you.

PANEL DISCUSSION

DR. PRICE: Thank you for those insightful comments. We will move to question one. And again, we may need to be adaptive, but we'll see how it goes.

Question one, regarding the scope of type one error probability simulations, should all mathematically possible parameter values for which the

drug is ineffective be included or only values that are in some sense clinically plausible? How is clinically plausible defined, agreed to?

So a couple reminders. Please lean in when responding. Please turn of your mics when not speaking. And we'll begin with Dr. Emerson.

DR. EMERSON: So I -- you know, it is impossible to do everything. But it is not impossible to use good statistical theory to look to say where the biggest problems lie and that that should be done to make certain that you do that.

The null space that was mentioned about saying, oh gee, you know, x is equal to y, but what equal to what, is not really as interesting to me personally as wondering about whether we're interested in a strong null or a weak null.

And that's very, very important to me, personally, is the idea that yes, if under the null do we really believe the treatment does nothing, a and therefore we should simulate from the exact same distribution, or we should simulate from distributions where if we're testing means, the means are equal, but

other things change, or if you're testing hazard ratios as returned from proportional hazards analyses.

If the average hazard ratio under the censoring distribution that you've used is the same, and, you know, they -- I focus a lot of this time-to-event. We do so many things in cancer based on time-to-event.

But you really have to worry about early differences versus late differences, and a lot of these making decisions very, very early when there might be an early difference, maybe in the wrong direction, maybe in the right direction. We really have to understand how that is there.

And also understanding how, you know, mean variance relationships then show up at under the weak null is very, very important. And so, it's really quite a big space, but trying to find some sort of smooth parameterization of this huge space becomes important.

And ultimately, you do have to stop that. You know, what's clinically plausible? That there needs to be something, but it's what's clinically plausible, not

just to the investigators at hand, but to the greater population of the people who have to be convinced.

DR. PRICE: Dr. Harrell?

DR. HARRELL: All right. First a couple of quick comments on Cyrus' presentation. The last method you presented for borrowing is really double-dipping, and is not a proper prior. So it's not a proper Bayesian analysis.

And then, the Bayesian and frequentist, when you have a non-informative prior, are only equal in a very special case. And that's where the sample size is fixed and there was exactly one look at the data. Otherwise, the Bayesian and frequentist are only the same if you don’t -- if you do not control the type one error with the frequentist approach.

But more to Scott's point, I think I'm glad he mentioned the strong and weak null. And the whole debate about whether we should be using null hypotheses needs to always be revisited. And there's a great debate in the philosophy of science about whether hypotheses lead to advancing science versus asking questions. And there's a difference in those two.

That debate's worth reading about.

And we have some famous statisticians, such as John Tukey and also Cohen has talked about this, that the null hypothesis is never true. So just get it out of your mind that there's exactly zero effect of a drug.

And so, when you're doing a simulation to use something that's artificial, it has real implications on what happens. And one way it's artificial is even if you believe the null hypothesis can be exactly true, which I do not, you have to also realize you're making another big assumption in the simulations which is you're entertaining that there is no possibility that the drug does worse than the control.

And so you're -- the option that delta is negative, the blood pressure actually gets higher with the new blood pressure drug, is not entertained. And so that's just one of the ramifications from doing simulations at delta equals zero.

DR. PRICE: Dr. Ashby?

DR. ASHBY: I mean, what's going through my head is which of these are fundamental questions and

which of these are kind of transient questions because we're feeling a way, and this is all new stuff.

I mean, I'm sort of thinking 10 years down the line, what you'd really want to see is probably some shared agreement of software. And it seems to me there's then two types of questions that you'd say, first of all, are these the right simulations. And you might have some agreement between regulators and sponsor about whether they're right.

And then secondly, whether they're technically correct, but rerunning them in different software. I think we should be looking for technical solutions too.

But you can't possibly go over all space. But I'm slightly worried that going down this line of questioning, I think we have to, sort of takes away from what was the most sensible thing to be doing. Is it really checking out the questions, the characteristics we need to understand? And I'm more concerned with getting the right set of questions.

So even when you start them at type one, it kind of forces you into certain sorts of designs and simulations, some of which are relevant, some of which

aren’t. So I think, you know, you almost want some interim guidance while we're finding our away. But I suspect the world in 10 years will look quite different.

DR. PRICE: Thank you. Dr. Goodman?

DR. GOODMAN: Sure. Thank you. Now, this is more a question. It's either a question or a comment. Very often when I see the simulations, they don’t always take into account the aspects of the design the clinician might have insight into.

And I think in particular I think of the temporal drift in the control group, which in combination with adaptive randomization sometimes -- I'll defer to Scott -- can introduce problems if you're adapting strongly and then all of a sudden your control group is doing better over time.

The other is clustering within sites. And so, my question is ae these -- very often that these clinical features of the study or of the procedure, the patients aren’t incorporated into the simulations. And the question is how to make sure that they are, that that input is there, because the clinicians don’t know

what questions to ask.

They can't look at the model and figure that out, and even know that it's missing. So it's really up to the simulators to be able to elicit those features that are not necessarily encoded in just the null hypothesis that would particularly interface with the design of the trial, particularly if it's adaptive. That would cause it to be misleading.

So this is sort of a question to those who do this for a living. It's also an issue for the FDA because they have to know how to ask these questions.

DR. PRICE: Any response before we move to Dr. Lewis? Dr. Berry?

DR. BERRY: Sure. I agree completely. And it's the -- to elicit that and simulate, for example, time variation and does the control vary. And it can be a problem if your control goes -- you randomize less and it was built up at a period where it's different. Could absolutely be a part of that.

And so, understanding from the design what are potential weaknesses or parts that could go away and the clinical possibility of them is the interaction

that has to happen whether it's ahead of time or once you've submitted and the FDA asks for such things.

The FDA may have further insight in other trials that are running, and they've seen things that then they could ask can you add this, can you add this to it.

One thing we've come up with on this front that's been a challenge in submitting is we don’t submit designs we're not running. So we submit only the design. Here's how it behaves. But we don’t show the work in which we decided that was the design, and we didn’t do this fixed trial. We didn’t do adaptive randomization because maybe it has this issue or that.

And so, that's a hard thing, and sometimes we're asked by the agency show us designs you didn’t decide so I can agree that this is better than those. It gets a bigger deal, and it can be confusing at times.

DR. EMERSON: Just, you know, a couple comments on that. You know, some of the temporal drift that you talk about conceptually could be dealt with with analytic methods, and therefore you do that,

although it raises small sample issues, which small sample issues have lots of things in terms of both heavier tails, in terms of they're caused by imbalance, caused by also mean variance relationships that cause greater problems, and the covariate adjustment with small samples. Then, that also is a bigger thing.

But, you know, some of those will pick out that you handle the temporal adjustment correctly in the adaptive randomization providing you've got blocked randomization and you adjust for that.

DR. PRICE: Dr. Lewis?

DR. LEWIS: So a couple points. With respect to the question actually posed, I think we're all agreeing that one should not try to build what I call statistical Maginot lines. That's protection against threats that don’t actually exist, such as the patient suddenly becoming immortal.

So I think the key is when we're communicating in teams, why it is we're willing to give up this false goal of analytic control. I think we need to point out that these design decisions have real implications. These are real tradeoffs.

If you limit your simulation for type one error control, if you will, to those things that are clinically plausible or at least worth worrying about even if you don’t find them that plausible, which is a broader category, that allows you to focus on the real threats to success of your program. And that's where our intellectual and computational effort should be focused.

I do thinks it's important that the clinicians or the scientific domain experts who view the simulations and give input on the advisability of the different designs learn to understand those things that threaten the success, as Dr. Goodman's pointing out, and this is a mutual learning experience.

And as Karen Price pointed out, you often get insights into the weaknesses of your designs through simulations that you wouldn’t have otherwise. And I think it's imperative that all of us work to suggest the kinds of questions be asked, like how sensitive is the design to temporal changes, to cluster effects and to those sorts of things.

But the main point I'd like to make here is

that there is a group that's ultimately going to be tasked with the oversight role when the trial is run, the data safety monitoring board or the DMC.

And one of the things that is critically important is that the DMC understands the range of assumptions over which the design's performance has been simulated and when what has actually happened falls outside of the domains that have been simulated.

Because the -- obviously no matter how much we think we know about a disease, occasionally things go terribly, horribly, differently. And the DMC has to understand when the design is performing or being asked to perform in a domain in which its behavior is well-understood or has been simulated. I should avoid that term, or when they need to worry that what they'd been told about its type one error control or other characteristics actually doesn’t apply anymore because the event rate is very different or the timed information is very different or the cluster effects are very different.

DR. PRICE: Dr. Irony?

DR. IRONY: I just wanted to talk about the

importance of working with the clinicians, and actually with the FDA reviewers.

You know, this is a pilot study. But as we get more and more experience with the simulations, as I got some working with Scott in the center for devices, you will know about these cases.

For instance, what happens when you vary the accrual rate? Now, that can be crucial. You know, in your first trial you maybe didn't think about that, but you will go, well, what happens when the control group starts to become better because the standard of care becomes better or that stats will become worse because you start to recruit patients that are worse off.

So all these things have to be worked in conjunction with the FDA because their reviewers will have experience as the clinical pilot is in industry. We will require more experience. So all these plausible scenarios and critical scenarios will be developed and can be developed in certain -- you know, with time.

DR. PRICE: So we'll have Dr. Mehta, followed by Dr. Bretz, and then we'll move to question two and

we'll have further discussion.

DR. MEHTA: Thank you. I think that the issue of borrowing data for your trial is a really important issue and should be thought through carefully.

You know, how -- I understand Frank Harrell's point about double-dipping. The issue is how are you going to not contaminate your actual trial, but when you bring in data, and this is actually what happens when sponsors submit their designs, at least to the center for devices, where this has been allowed.

So I presented a method that was proposed by the MDIC in a public meeting. I haven’t developed this method. But you have a choice. You can say, well, there are -- FDA sometimes says we give you a fixed discount. You can take 30 percent. Either take it or leave it, That's the kind of attitude.

Now, then that places the sponsor at a risk, and it also places the regulator at a risk. It places a sponsor at the risk because if the new drug is positive and the historical, when it becomes available, is negative, he's going to force to take 30 percent to pull down his affect.

It's a risk to the regulator because if the new drug is ineffective and the historical drug is positive, that's going to boost up the delta. So there has to be some kind of a dynamic way to allow borrowing or else you should say we don't allow -- we don't allow borrowing.

And maybe there are better ways to decide without introducing biases. But I think that's something that needs to be thought through and discussed.

DR. BRETZ: I guess I'm not sure what the difference between question one and two are. So I'm trying to address both.

And I think I'm hearing multiple conversations going on. And that makes the conversations particularly interesting, I guess. But starting this what we mean by a null space, Frank, if you mentioned that you don't believe in the null hypothesis, you're referring to two-sided null hypothesis. But what about one-sided. And I would only think about one-sided null hypothesis.

To me, that is important because if you think

about the null space, and then from a multiple testing point of view in the closed test procedure, the strong versus weak element control, then it kicks in in a multi-arm trial, and you would have to think about simulating all combinations of effective and non-effective arms and the effect sizes.

If you think about an enrichment design, all combinations of treatment effects and the subgroup and the overall population. And if you think about adaptation routes, depending on surrogate safety, secondary endpoints, all these effect sizes would have be to included if you think about the strong type one error rate control. So this is one comment about the null space.

Then I'm hearing another conversation which is quote natural, Bayes versus frequentist, and maybe overlaid or underlying that a type one error rate control is necessary at all. And I don’t think we have the discussion today here. But that's why I like the initial comment -- was it from you John or Lisa, that simulation is not Bayesian versus frequentist.

It's -- and I like what Scott said, it's an

integration problem and I agree fully with that.

And if you think about just a quantile of a standard normal distributions, the 1.96, we get it by Monte Carlo. We can also get it by simulation, and no one would ever have any problems because it's a very defined simulation pattern, right, problem. It's a well-defined integral, and we can solve it efficiently or in this case inefficiently with Monte Carlo simulations.

If you think of it by variate probability where you have a nuisance, like the correlation parameter, it is still tractable. It is a well-defined integration problem because now you can go through the grid different correlation values and you can kind of understand what is your least favorable configuration.

That's quite possible numerically if you use a very fine grid. But then, I think transparency is an important topic when you -- when it gets more complicated designs and more advanced problems. And if you are thinking about how to define, if you -- I'm hearing about accrual rate and historical data, control rate changing, what is then the integral.

Probably you cannot write down the integral in a closed form anymore. So you can simulate the trials, and it remains somewhat an integration problem. But it's somewhat a black box what actually your integration problem is.

So this transparency I think is important, and if -- I think one step ahead on -- if you think about -- which is not a topic directly related to simulations, but if you think about transparency of analytical calculations, if you think about this deep neural networks where you have hundreds of layers and thousands of parameters, it becomes incomprehensible what such an algorithm is doing.

So where is the boundary of transparency versus then accepting algorithms for which you can't follow anymore what's doing. So I see there's some relationship. What's the necessary degree of transparency in terms of the assumptions, of the values being simulated or power of the simulation program.

If you want to evaluate it, I think these are all interesting questions, I think very difficult questions from as far as I'm concerned.

DR. PRICE: So we'll move to question two. And I think question two has two parts.

How should error rate simulations be conducted when formally borrowing prior information such as in a Bayesian framework, and what does type one error mean in this setting? Should we consider other error rates instead?

And just looking at our time, if someone also wants to respond to question three, what are some practical suggestions for implementing trial simulations. Feel free, and I see quite a few cards. So Dr. Price?

DR. K. PRICE: Thank you. So I am going to sort of combine two and three I think in this response. So foundationally, I think we can just think about error rate simulations generally, and how best practices, and then talk about some additional components that might be relevant in a Bayesian framework, some additional pieces that would need to be included.

So in terms of best practices, obviously you need to have a really good simulation strategy that's

laid out ahead of time clearly articulating the objectives, how will virtual patients be generated, methods for ensuring that how -- what you intended to generate was actually what was generated, methods for how you analyze the data, things around have your simulations converged, and there are graphical ways that you can do this.

I would argue that one area we could improve is our ability to graphically think and look at summaries of simulations. Maybe there are some interactive tools that could be developed to help look at the simulation properties.

And then of course with -- if you have a prior, what is that choice of prior, maybe exploring some alternative options to show prior posterior sensitivity, looking at convergence of the prior and things of this nature.

So I think that there are -- there are some references that can be included here where people have talked about best practices, some of these coming from the adaptive design working group years ago, some coming more recently from Bayesian scientific working

group and others as well. So I think building on those would be really important.

The one last suggestion that I think is very helpful, and I've heard some others mention this, is sometimes just looking at examples, like specific clinical trials that have been simulated, and how this plays out.

You know that's something, Scott, you've shown in the past. And so, those types of things, always looking at maybe certain edge cases or where weird things happen so that we can have a conversation about why that happened. Those types of things seem -- would be good. Thank you.

DR. PRICE: Dr. Berry?

DR. BERRY: So this is a great question, and I swear if -- in a weird way, I enjoyed John struggling with this question and physically grappling with it because it's really in a situation -- and I understand Cyrus' point about borrowing information.

But in a situation where is deemed reasonable, the data is deemed reasonable, that this is a good thing to do for drug device biologic development,

you've already started with a leg up. There's a reason to bring the data in, to go back and simulate from a null and say, aha, the chance of approving this drug is 11 percent because I'm using data that already gives it a leg up. It's a weird type one error because we've already said we actually don't believe it's very likely, you know, that even within some small delta of that.

Now you can say okay, well let's go back to before that data was observed and then simulate that data being generated and then go and what about that whole process? But it's hard to go back and retrospectively generate what data could've been there.

If you're in a situation where you believe it was cherry-picked, I don't think anybody here thinks it's reasonable to use that data, that we're doing multiplicities and bad science, and I don't think anybody would be in that situation.

There are more complicated scenarios where you're in an antibiotic situation or borrowing from a previous trial and the control. You inflate type one error when you borrow information even on the control.

Every single arm study has type one error of one in different situations. But we allow them to happen because we have a certain comfort with them.

So I think we should simulate different, quote, unquote, "type ones" and understand the unlikeliness of them or the behavior of that, but not live under this framework that you have to be 0.025 or less.

Otherwise, you would've never borrowed the data in the first place, and understanding what is reasonable in that scenario is a very common thing. And Telba made reference to this. We've done this with Telba several times before. And it changes the conversation of the balancing of these errors as opposed to the control at a certain level.

DR. PRICE: So we'll move to Dr. Lee and then Dr. Emerson, and then we'll open the floor up to the audience.

DR. LEE: Thank you. So I'd like to give some comment on question number three, what are the practical suggestion for implementing trial simulations?

Well very simply, make the tools available, okay, to people, and then there are some -- you know, make sure that it's easy to use, and, you know, it's well-documented and it's reproducible. And actually Karen brought up a very good point, is that you need to provide some graphical assessment so that people can see, you know, how things work.

And for me, the eye-opening thing is when I start to learn about how Bayesian method works and how simulation, you know, can be conducted. People asked me -- some people asked me during the break. You know, I mentioned briefly about the software developer at MD Anderson. And let me just say it again. If you Google MD Anderson software, download our software online, okay, or even go to trialdesign.org, you know, you can get those freely available. I don't know if it's out of the line or not, you know?

If there's Internet, you can just type trialdesign.org and there are many tools available, you know, from learning the Bayesian method for both binary endpoint, continuous endpoint and time-to-event endpoint or design trials, give you the stopping

boundary, you know, of say Bayesian toxicity monitoring or efficacy monitoring and then give you operating characteristics. And even you can actually run the trial using some of these method, and you can actually also in some cases, after you design the trial, it can give you the template that provide the statistic consideration section in the protocol

So anyway, my passion in academia is that try to make the tools freely available to people so that we can learn and we can improve, or in some cases, you know, we can debate the choice of the prior and what's the impact of the prior, et cetera. But it's all there, okay?

So as long as the tool, you know, is available, and is reproducible, and we can -- or, you know, agree -- you know in some case, we can agree to disagree, right? And but again, the thing is make it transparent, make it -- you know, kind of useful, reproducible, then we can make improvement upon it. Yeah.

DR. PRICE: Dr. Emerson, did you have one final comment?

DR. EMERSON: So, you know, in talking about how we implement these simulations -- and Scott made reference to the guidance's, you know, less well-understood methods.

And I guess some of my contention is actually what's less well-understood is actually the statistical methods interacting with the adaptive methods rather than just the adaptive methods themselves, is that there's less room to not understand the statistical methods you're using and still have it be valid.

And the idea that particularly our methods are often presuming such things as we're using sufficient statistics at all times in their adjustment and there's no other information in the data.

And, you know, Roger made a very good point of saying, well, simulations -- we have to really -- it has to be pre-specified for us to simulate this. But to then take these simulations and believe that they'd ever carry forward to fully adaptive, not pre-specified is very dangerous because there can be additional information that's never been considered in that setting.

In simulating this, one question particularly for simulating Bayesian methods, you know, I don't know whether I'm -- you know, after Bayes estimators or mini max estimators sometimes, and realize that as you take a prior, you're averaging over lots of individual alternatives.

And when is it better to maybe sort of do a tipping point analysis of saying here is a fixed alternative that we behave less well in, and then later you can average in to say yeah, but I don’t really believe that that much, rather than always averaging over those. So it has some issues.

And then, as Cyrus pointed out, you know, to be really certain what the type one error was, it can be prohibitive. And so the question is how do you decide what's there. And I'll just say what I tend to do is I take our standard frequentist methods that nobody blinks at and say what really is the type one error in those situations and how high can it go.

And recognize that on the type one error, if I take my simplistic world of a binary decision space and a binary parameter space, the Bayes factor is the power

divided by the type one error. It's multiplicative.

And so, what sort of multiplicative increase in our type one error will be accept? And if you just take the standard chi-squared statistic that we take all the time and don’t worry about it, it's -- we don’t have to pick out 0.026 versus 0.025.

You know, we'd feel fairly comfortable with 0.027, 0.028 and adopt those standards and just say yeah, it's -- we don’t need that precision. We just need to be a standard that people can operate on and define what the level of precision we need to be certain we're okay.

DR. PRICE: So we have time for comments or questions from the audience. Dr. Louis?

AUDIENCE Q&A

QUESTION: Just a comment. Tom Louis, from Hopkins. Sitting here listening, I really got impressed by the power of simulation really as a catalyst. Even if you were to never simulate, the conversation that's happened for the last hour or so just wouldn’t happen if you couldn’t imagine simulating and have a discussion range widely about things that we

couldn’t do analytically.

And so, I'm not saying we shouldn’t simulate. We absolutely should. But it's a little bit like the bootstrap. Discussing what you would do is almost as informative as doing it. And so, I really think simulation's role in getting a conversation going in a way that's hopefully biologically and statistically complex is important.

And then another comment, just sitting here reminded me of a conversation I had with Brad Efron about 10 years ago when he said Bayesian's get all the glory, but frequentists do the hard work. And the simulation exercise is pretty much a frequentist activity.

QUESTION: Jonathan Smith, Adaptive Plus. I want to -- first of all, let me say that my comments are really in relation to frequentist designs where there's no adaptive randomization and we are saying we have to control -- have strong control for type one error.

The first comment is that there are many more situations where we could actually use a numerical

integration rather than have to use simulation. And that's quite often plausible, up to maybe six dimensions, perhaps even more.

Then if you are using simulations, I think that it -- this may be a couple of strategies that are useful to try and get to the maximum.

The first is perhaps start off with a fairly wide grid across each of your factors. Try and find the area where -- or areas where the maximum type one error lies. Perhaps use Scott's idea of modeling to try and hone in on that region. And then, once you've got closer to that region, then run your simulations, maybe with 10 million, 100 million.

Another option would be to consider maybe you've got five factors that impact your type one error. Rather than looking at a complete factorial approach to your grids, maybe instead of -- I don't know, if you've got five levels of each of your factors, instead of looking at that, you could start out with a factual factorial, maybe 125th of the 55 factorial, which is only 25 different cases. And that's going to give you a lot of information. So

those were just a couple of comments. Could I make one more, or --

DR. PRICE: Briefly.

QUESTION: Sorry. One more comment, a slightly different topic, and that's the follow-up from Frank's comment about whether we need to consider type one error under situations where an effect size could be in the opposite direction.

And I think that in the subgroup situation, that maybe is the one place where that often comes up, and we've certainly seen situations with -- where a biomarker negative subpopulation does have a negative treatment effect. Okay. Thank you.

DR. PRICE: Since we do not have any additional questions from the audience, I'll return to the panel because we have a couple more minutes. Dr. Mehta, I had actually skipped you because we were going to the audience. Did you have a comment to make?

DR. MEHTA: No.

DR. PRICE: No? And does anyone else on the panel? Dr. Harrell?

DR. HARRELL: I think it was Scott that

addressed this about having a fixed, unknown parameter you're simulating from. Was that you that just hit on that a few minutes ago?

DR. EMERSON: Sure.

DR. HARRELL: Well, I don't know which Scott it was.

DR. PRICE: Scott.

DR. HARRELL: It was one of the Scotts, maybe.

DR. BERRY: What's your prior probability was he or me?

DR. HARRELL: Yeah. So the thing I wanted to mention is the -- when I think about that issue, I think about what is the purpose of Bayesian inference, is to -- is to take whatever the world is throwing at you and be able to gain knowledge about what that thing was.

So you're trying to -- you're trying to make a prediction or an estimate. You're trying to recover what was not known and to make it more known. So in that way of thinking, the way you would do a simulation is you would have a whole variety of unknown values and you would ask the Bayesian procedure to what extent

could you recover those unknown values, whatever they were.

And that's actually much different than doing a simulation that sets the unknown value to a constant and finds out some operating characteristic, Bayesian or frequentist. If you're trying to recover that constant, you could never recover the constant.

You never would, no matter -- unless your sample size was infinite. But those two ways of simulating are much different, and I -- the Bayesian one to me is more natural because I'm trying to recover whatever it was that was generating the data.

DR. EMERSON: So I think you are responding to me, and we'll discuss whether it was well or not.

The problem that I have is so often I face post hoc interpretations of data that are radically different from what the interpretations had been beforehand. And it's impossible to, you know, in any prior anticipate something that you never expected to be the case.

And so, all I was advocating is that finding the limits of, you know, point priors, if you will, as

well as what an average effect is, is something worthwhile because as I was just remarking in a sidebar to Frank, is that I have been on DSMBs as many times where the treatment effect turned out to look harmful at a magnitude that they were looking for it to be beneficial as I have found that have materially turned out to be that beneficial.

And nobody started the trial thinking it was harmful. And if you elicited priors forever, nobody would've given you that prior. And so, by trying to find those limits, think of it as a tipping point analysis. That's additional information, and then at the end you can say how unlikely you think it is.

DR. HARRELL: Just one slight rejoinder. I think most people would use a prior that would allow for that to happen.

DR. EMERSON: But populations of priors are what matter. So I don’t like consensus priors. And in fact my -- it is a polling. We're trying to go our to the population of scientists and see how many we've convinced. We're not trying to see if we've only convinced the average person.

And so, it is relevant to me also to say I don't have to prove things to everybody. But have I proved it to 95 percent of scientists, 90 percent of scientists, and that also helps.

DR. PRICE: Dr. Permutt?

DR. PERMUTT: I want to go back to this question of how big a space and can we limit ourselves to clinically plausible values.

And you all design trials, and you've addressed it appropriately from a point of view of trial design. But let me ask you a question as a regulator that maybe you can advise me on.

After the trial is done, we have a lot more information about the nuisance parameters than when we started. As a regulator, can I confine myself to the values of the nuisance parameters that are plausible a posteriori, and not necessarily worry about possible configurations of parameters that have by now already been more or less ruled out?

DR. PRICE: Dr. Berry?

DR. BERRY: So I brought this idea up, and I'm sitting next to Dr. Meurer, and we have a trial, a 10-

arm cooling trial after cardiac arrest. We don't know how long to cool, so 10 different arms, inverted u-dose response, Bayesian primary analysis of whether or not cooling actually helps, even though we don't have a zero cooling arm. Really quite hard to analytically understand the characteristics.

It depends on the distribution of modified rank and scores. Are they more towards zero? Are they more towards six? We simulated a wide range of these, and then we created a complete plan prospectively that we submitted to the agency. This was the center for devices.

This is how we're going to simulate it after the fact. We're going to condition on what we observed for these nuisance parameters and re-simulate it, because we had to do this grid search for type one error.

And we're going to plugin those values, re-simulate it, and what we're going to bootstrap the actual values of the patients, ignoring the arm that they came from, to just further explore this type one error.

And we're not going to then adjust the final criterion, but we might report that the trial had 2.54 percent type one error. We believe we've covered the space. But this gives us additional after the fact simulation of type one error where we plug in nuisance parameters that are now known.

DR. EMERSON: Or at least now estimated. So that they -- and you know, realize -- and again, when you look at standards and what we do -- I mean, this is standard sort of maximum likelihood idea of you take the estimate of the nuisance parameters and then you just look at what the variability is conditional on that.

Now, we know however that that doesn’t work entirely, and we know that bootstrapping can't work, but maybe double bootstrap can. And we have to worry about the fact that always our alternatives are counterfactuals, right?

We're saying gee, I observed a difference of 10 in the means. Could this data come from where -- when the difference was really zero? Well, if I was that far off on the mean, then how far off was I on the

variance? And the central limit theorem says concentrate on the first two moments, because we blow apart all other information in the data.

And so, thinking about what the variances might do under these various alternatives. So the most I can come back to is sort of the thing of saying well we're sort of back to standards here, aren't we? We don't ever know the exact truth, and people can -- a mean variance relationship is never identifiable because it is a counterfactual.

It's saying that the data -- you know, if our data is wrong in some sense, then it can be wrong on what that variance would be under something else. But we can come up with standards, and we can come up with standards that may not be perfect for the problem at hand but that are on average okay.

And so, I do tend to worry about -- again, it's all just do tipping point analyses with everything. But I look to say how bad does it have to be? How bad does our current data have to be before I, you know, can't trust this at all?

And then you sort of go forward. The more

that the nuisance parameters are not totally nuisance and are really more correlated with what your effects are, the more trouble you're in.

But the other comment is focusing on what's plausible, there's all definitions of what's clinically plausible. So while I just did both, you might tell me that you don't think it's plausible that the treatment's harmful, whereas I'll always say oh, yes it is.

But I'm not usually going to put hazard functions that are step functions. They're going to be fairly continuous, and I'll restrict myself to continuous hazard functions on my data.

DR. PRICE: So we're going to move into our break. I encourage panelists who I didn't get to, to have offline discussions during the break, and we'll reconvene in --

MS. BENT: 15 minutes.

DR. PRICE: -- at 2:45.

(Whereupon, the foregoing went off the record at 2:35 p.m., and went back on the record at 2:49 p.m.)

DR. PRICE: So we will move into our final session for today. We are being cognizant of the weather, so we will do our best to end no later than 4:15, and if we are told that we need to push that up, we will.

So first I would like to have my two colleagues, who are joining this panel to introduce themselves, starting with Dr. Johnson.

DR. JOHNSON: Laura Lee Johnson, Office of Biostatistics in CDER.

MS. KRAUS: Stefanie Kraus. I'm a regulatory counsel with the Office of Regulatory Policy in CDER.

SESSION IV: COMPLEX INNOVATIVE DESIGN PILOT PROGRAM PRESENTATION

DR. PRICE: Thank you. So the final session will focus on the complex innovative design's pilot program. The pilot program will have many goals, including an increased awareness of the value of complex innovative designs in a wide range of therapeutic areas and increased learning and sharing, both internally at FDA and externally to FDA.

I could name other goals, but I think we're

all gathered today with the anticipation that use of complex innovative designs will further enhance drug development, which ultimately translates to benefit to patients, and our overall public health.

The purpose of this session is to engage in discussion and obtain input on the pilot program from a broad range of stakeholders. I will only give a high level overview of the pilot program as found in PDUFA VI.

Today's discussion will be used to further inform our thinking on various aspects of the program, as well as implementation of the program.

Next steps for the FDA will include a Federal Register notice announcement of the pilot program as well as development of a CID website, CID being complex innovative designs, with all pertinent information.

So under PDUFA VI, the FDA will conduct a number of activities that confirm our commitment to advancing the use of complex innovative designs.

The activities include development of staff capacity, conducting a pilot program, again a focus on this session, convening a public workshop, such as

today's meeting, publishing draft guidance and developing or revising relevant manuals of policies or procedures and standard operating policies and procedures.

The pilot program is designed for highly innovative trial designs for which analytically derived properties may not be feasible. And simulations are needed to determine trial operating characteristics.

Sponsors may submit designs to the program, and those selected will have the opportunity for increased engagement with regulatory staff through two meetings.

FDA will select up to two proposals per quarter. The agency will use the designs as case studies for continuing education and information sharing.

In terms of the increased interactions, FDA will grant a pair of meetings consisting of an initial and follow-up meeting on the same design to occur over a span over approximately 120 days. The meetings will be led by the statistical review components within CDER and CBER, but will be multidisciplinary.

So there are a number of elements of the pilot program that we're currently working through and we welcome discussion on these points today from the panel.

The elements include eligibility or selection criteria for the program; timelines, for example, timelines for submission of proposals, for selection of proposals, for review of proposals and for feedback; submission expectations, for example, what are some of the key elements for the proposal versus the meeting package and are they the same; disclosure, so a unique aspect of this CID pilot program is the FDA's ability to publicly discuss example designs to provide clarity upon the acceptance.

Before granting the initial meeting, FDA and the sponsor will agree on information that may be shared publicly; communication, this could include a communication strategy between FDA and sponsors of selected proposals, as well as strategies to inform the broader audience.

So before I go to our two reactants, I will go over the discussion points. The FDA will select two

proposals quarterly for entry into the pilot program. The proposals will need to capture sufficient details to facilitate an understanding of the design analysis.

Discuss specific elements of the design and analysis that are important for the initial proposal. Discuss types of trial designs that should be prioritized for selection into the pilot program and discuss factors that might inhibit or encourage submissions for the program.

So I would like to ask our first discussant to provide remarks, and that will be Dr. Lieberman.

DISCUSSION

DR. LIEBERMAN: Thank you very much. So I think we've heard a lot here about flexibility and creativity. And the pilot will definitely require some flexibility, creativity, mutual understanding, patience on especially the sponsor site and transparency in decision-making, both on the FDA side and the sponsor side, as well as communication, very clear and open communication constantly.

The thing -- maybe before we go to some of the timelines, and I leave to my other speaker some of the

other topics, but before I go to some of the timelines and submission materials, I think it's important to emphasize that even though this will be statistical meetings, it would be nice to understand the role of the clinical reviewers, and so their acceptance of the study designs because it's not just statistics, but it's the endpoints. It's the selection of the populations. And all of these factors will influence the simulations, right?

And then, if it's a study involving diagnostics, will CDRH be involved? Will they be part of the meetings, be able to at last say yes, the data that will be generated from this study will also support a filing for a diagnostic. So I think all these will be important.

And then, we talk about timelines. You want to think of the flexibility because every sponsor has their set of own metrics of what is fast and how fast they want to move and what are the risks they're willing to take, and different types or kinds of developments.

So sort of it's hard to say that one size will

fit all. And what I mean by this, some sponsor may say well, we like to as fast as possible go forward after the end of phase meeting. So when we reach our agreement, one month or two, we'll ready with our protocol we're running.

This may be all very well. But then where do you accommodate the 120 days between the meetings and the timelines? Other sponsors might say okay, we're not rushing. We're going to take more time.

So if we're going to look at this, somebody says yes, we're going to apply like six months, way before the end of phase meeting, and start the whole process. But will there be enough information at endpoints?

Will there be enough information on the populations you want to select? Will there be enough information on the safety data to make the right decision about all the simulations and the right study designs?

Well, maybe the statistics will be hemmed out. But then if the parameters change, it's a risk to the sponsor if things change. So then there's the

flipside, let's wait. Let's wait with all the materials. When we're really ready, we submit more or less at the same time as we ask for the end of Phase II meeting.

Okay, well that means that the timelines will push us and we're going to have to look at -- we are going to be far after what we normally'd like to do, two months after end of phase meeting, ready to go with a final protocol. So I know these will be the considerations that the sponsor will take.

So sort of when you think about it in summarizing, I think it will be nice to think that there's going to be a window that the sponsor could submit their information, and just don’t be strict about it. Oh, it has to be right at the time of the package for end of Phase II meeting, or it has to be x weeks before or after, but put a window around it.

When we looked at it through the bio company, we said, you know, it could be anywhere between 60 day ahead to almost -- to a few days beforehand. And then that will sort of define what is really in the submission.

If it's close to the package, then the sponsor can reference a lot of the information in the package because that will help the clinical reviewers understand the study better.

If it's way before the package is ready, then maybe they need to be more information, not just about the study design, not just about the statistics, but also a little bit within the context of the whole development program to allow more sort of a better perspective on the material.

It would be nice to understand, you know, what is the timeline for getting response to the application. Could it be a month to 45 days, especially if there has to be negotiations about the disclosure of relevant material?

So sort of say, you know, after two weeks if there is no issues about disclosure, there's just -- the sponsor will know. You will know the answer in an additional two weeks. But if we have to discuss disclosure, it might take another month.

The other thing is sort of, okay, after the acceptance, what is the time? What is -- and part of

it is can we preschedule the meetings at risk ahead of time or what is the feasibility of scheduling a meeting once the acceptance happened? Could it happen within a month? Is it even feasible, both for the agency and the sponsor? It may not be.

So maybe there's a way of sort of thinking ahead of time of how do we don't have these wide spaces between the times we make a decision. Oh, now it's going to take us two months to schedule a meeting. Can we be flexible of how we think about this?

Then the sort of -- the next step would be, you know, the meeting between the two statistical meetings. I would expect that the 120 days is really for the statisticians at the FDA to run the simulations.

So if it's a heavy simulation project, yes. But if it's not, is there any feasibility to sort of making the meeting 90 days or less, so negotiating that, depending on the scope of what needs to happen between those two meetings.

And then, how do we sort of, you know, communicate through the whole process? It would be

definitely good both on the sponsor side to have, especially if it is heavy statistics, if there is interesting topics, somebody from the sponsor being really the key contact person, but also have someone on the FDA being the contact person for the study so there is clear communication.

I would also envision that the packet, that the submission is different than the sort of meeting -- material meeting presentations. But that maybe does not have to be sent like weeks ahead, but maybe just two, three days before the meeting. And that would include, I assume, all the details of the statistics, all the simulations information that will be then discussed at the first meeting.

And the same thing if there is a review of the statistical sort of design on the agency side prior to the second meeting, it would be good for the sponsor to get sort of overview of what happened, at least few days before the second meeting so that they can look at that react, versus coming to the meeting and here are all the problems that we found with your design, right?

So sort of allow that flow of information. So

that's why that 120 days is probably critical to allow for this to happen. So it might be hard to do. And of course, I would envision that any documentation, any simulation results that that would be submitted in the standard form.

I believe there were some of -- in the previous session sort of ideas of what the reports could entail, so that would be another thing. And I think that I might stop here and let me friend continue.

DR. ZHONG: First, I would like thank the FDA and our colleagues from the industry and also from academia who come here to have this discussion.

The FDA this initiated this innovative clinical trial pilot program. Actually shed a light on the new era of drug development and gave a lot of confidence to the sponsors.

As a lot of people pointed out early on, there's a lot of misconception either in the industry or somewhere. It may or may not be in the agency, but somewhere. Maybe because of lack of issue with the opportunity for innovation, right?

So first I would like to make a statement there, thank you. And secondly, I have to be fully -- I have to fully disclosed that Gracie and I are coordinated, and that's why she did not touch all the points there, and she -- we can spell out on what we're going to talk about.

And third, I mean, I agree to you -- make a transparency that the points I going to talk about actually coming from some consensus from the bio group, also some from the pharma workgroup as well. It's not purely my personal belief. It's some consensus form there.

So first let me touch on the entry criteria. For the entry criteria, I mean we completely agree with agency that I mean the whole purpose here -- actually the pilot program is to promote public learning, like in -- and promote statistical innovation and all innovation in health.

And the type of designs that Lisa early on, presented actually kind of hit the nail on the head. That's the kind of design that we need to see, okay? Even though as Lisa point out, maybe statistically it

may not, so in a way, but regulatory, it is innovative. That's the point I make early on. That's the opportunity for us to bring more awareness on the openness at the agency and to the whole industry to kind of bring the innovation to product development.

And also, some speakers early on also said let's focus on those kind of designs that used to be labeled as not so well-understood designs. Yes, and we're all in agreement there.

And those innovative design that the agency or the industry have less experience with, even though it could have been used in one or two indications, or by one or two other companies or divisions. But just because of misconception there, it's good to take those trials into the pilot program.

And then, the third is focus on candidate on pre-clinical trial stage. People try to define it as have they impact on drug application, like the DRA, NDA, like focus on those first. We all know that a lot of innovative designs have been used in early trial designs, and FDA never discouraged those. We all understood that. But I mean, the issue is could we use

them to establish the evidence to inform regulatory decision on drug approval? So this is area that we would like to focus on.

And lastly, the entry criteria should be like a -- I mean, like sponsors of the applications should be amenable for public discussion, to encourage public learning, learning at the agency, at industry and others in the statistic community as well.

And of course, talking about disclosure, it's not like we're going to disclose everything. I mean, I think the feedback from the companies -- I mean, is that what we should consider is that -- I mean, is only disclose information that is necessary to promote public learning instead of disclose everything, right?

I mean, we all heard from agency that, well, disclosure must be a negotiation between the sponsors and the agency, and then the -- but when we talk about the -- there needs to be a certain element that is to be disclosed, right?

Most companies are more open to disclose on a design element, such that mostly will likely to be statistical aspect of those kind of elements there.

So when we disclose things like product, please do not disclose those things. You can use drug A, drug B instead of exactly which product is, or even a disease.

A lot of companies are not comfortable with disclose which disease the company is targeting on. So those kind of information is not necessary to say. But I'd also like endpoints. Endpoints, you don't have to disclose the definition endpoints, but you can disclose if it's confused endpoint, ordinal endpoint or like cardinal endpoints. You can disclose these information like that.

And then in terms of statistics, like sample sized or maximum sample or expected sample size, like the power determination. Maybe we don't talk about power. But the false positive rate or false negative rate, all of those things. I mean, and then the simulation, I mean, we talk a lot about simulation. And the simulation is a thing that we all need to learn.

And in terms of simulation, I mean we're hoping that we can disclose as much as possible, but

has to be compiled with sensitivity and confidentiality of the companies. So that's about the disclosure. I mean there's a -- I mean the bio group actually submit a recommendation on the disclosure. But I'm not going to go through the whole thing, but I picked up the highlight there on it.

And then for the communication -- this communication is also something that's very critical for the success of the pilot program and for the encouragement of the companies, sponsors to participate in the pilot program as well.

So we kind of think that this would be good if the agency could identify certain elements, working with different working groups and sponsors to certain elements to disclose on a website, some -- like a website or in a certain form so that, I mean, all the companies who wanted to participate in the pilot program will have the opportunity to know what the status is.

I make one suggestion is that in order to promote public learning, I mean, the entry criteria there that we would like to look into all different

types disease area and still focus on one disease area, focus on looking all the different disease area, and also don’t focus on only one trial design type, but focus on different product design choice.

Yeah. So that's what we'd like to have, and I also want to point out, I mean, the bio group actually have a list of the elements that -- who I recommend to the agency that we would like -- usually the companies in consensus would be able to discuss on like what kind of information can we disclose on a website or some public forums, allow the transparency to the public.

PANEL DISCUSSION

DR. PRICE: Thank you. There was a lot in both of those comments. We are going to open the discussion up to the broader panel, beginning with Dr. Lewis.

DR. LEWIS: Okay. Thank you. A couple of just comments about the characteristics of the selection process are things that one might want to consider. One has to do with the diversity of sponsors.

Obviously it's important that traditional for-

profit sponsors be represented in the program. But as was mentioned earlier by Dr. LaVange and others, some of the work that we're seeing in master protocols is driven by things like patient advocacy groups.

And one of the things that the agency may be able to do to lower barriers, to research efforts in some diseases, and the multidrug-resistant bacterial pathogens is one that is near and dear to our heart, is to support the design efforts of these patient advocacy or other nontraditional sponsors to develop protocols or master protocols that have some indication that they would be acceptable to support regulatory decision-making because that's a -- the uncertainty in that regard is a barrier to sponsors then being willing to submit their compounds for testing in those platforms.

A second has to do with the likely nature and complexity of the simulation code that would ultimately be required to validate the design. One of the questions we get asked very frequently is do you have examples of publicly available code that people can run to really learn what it takes to write a simulation that is realistic enough to inform decision-making.

And I think there's unfortunately relatively few examples of that in the public domain. So within the diversity of proposals that would need to be represented in the program, it would be nice that if there was at least some in which when we anticipated that the code be able to be publically released and would be written in R or some other widely available format so that we could help to develop expertise in the broader community to do those kinds of simulation.

And then the last, which I'm afraid is maybe a little contradictor to the comments of my colleague, is the importance of concreteness. When examples are anonymized, they lose a lot of the power that they have. A lot of people are looking for very specific examples of something that was a successful design effort.

And so, I would hope that at least a subset of the proposals that are accepted into the program are developed in a context in which they can be as concrete as possible with respect to the clinical disease endpoint and other considerations because this is an area of clinical trial design where the -- those

clinical details are supposed to inform the design because the design is intended to be customized to those details.

So when you anonymize them, you break the link between the very information that is intended to inform the design and the design itself, and that will lessen the value of the examples.

DR. PRICE: Thank you. Dr. Chan?

DR. CHAN: Yeah. Very good comments provided in some of these discussions. And I just wanted to add couple things. One is that related to the elements obviously we talk about some clear layout or simulation plans.

And I think part of that is how having a clear layout and the simulation plan under code to allow whether it is FDA reviewer or third party actually to do a peer review, or even better, to do a validation of the code. So those definitely needs to be part of the elements.

But then in terms of the study design, I was thinking whether FDA actually also would willing to entertain maybe just on the program level, not just on

the study per say, because sometimes you need -- actually when you're thinking about a drug development program, there's several aspects that actually from one study or another study, they involve a lot of simulation or trying to do the monitoring to help select the dose and things like that.

So I wonder whether FDA would be open to consider sort of a serial study as one -- or a program, how that pan out.

And the third comment or question is in terms of communication time. I notice to you -- submit a program and FDA have an interaction review, but for the transparency -- for the public to understand, I wonder how long it would take for the FDA to actually share the study design, even though the study may still be ongoing.

But there's certain elements can start to be shared throughout the community so that the next batch of the program can be balanced along those lines.

DR. PRICE: So I don't know that we have answers to those questions. But they're definitely points that we will consider. Dr. LaVange?

DR. LAVANGE: So I'm not exactly answering this question, but I just wanted to make a few general comments about the program as we envision it.

So first, I wanted to give Estelle Russek-Cohen out there in the audience some credit because she was really the brains behind this pilot program during the PDUFA negotiations, and is particularly the part about being able to talk openly about the designs so that better information about what FDA will accept or not accept you know is out there.

Second, we'll -- may get into this today, but it'll be better explained when the Federal Register notice comes out announcing the pilot. But we did envision at the time, and probably FDA still does, that there would be different you know -- the disclosure wouldn’t be the same for every design.

So it may be that nothing about the drug or the sponsor or even the disease needs to be disclosed. It may be just the elements of an adaption, or something else.

On the other hand, if there is use of a patient registry, then you'll have to disclose

something about the disease. But you know this would not be just opening up the entire development program, or even the protocol.

It's not a full disclosure or protocol. It's really just being able -- I mean put yourself in FDA's shoes and think about what would be most beneficial to the world of directed element for us to disclose.

Well maybe it's the fact that we are seriously considering a design with this particular adaptation at these particular times, based on this information, these decision-making criteria, this level of evidence, or this way to simulate the operating characteristics and so forth.

So you know, it's not full disclosure of everything, and I won't say anything else, Stefanie. But I think that's important that this is a -- what would be disclosed or what needs to would be on a case-by-case basis, and we're not talking about in most cases opening it all up.

And then, you know, third, I think it's very important, and somebody has said this already today it's up to the sponsor. If they -- they don’t have to

partake in the pilot, and that doesn’t mean they can't propose something innovative. So you can still propose a very innovative design and choose not to go through the pilot.

The idea of going through the pilot is you get a little bit more regulatory interaction, hopefully you know pretty fast, which is -- Gracie made that a very important point. And you're contributing to the body of knowledge, which advances science. So you know it's a good thing to do, so just wanted to make those points.

DR. PRICE: Thank you. Dr. Ashby?

DR. ASHBY: I think I'm probably answering point two more than point one. Regarding -- so I think the first thing is that you need to be careful about the range you accept, because once you start going public and talking about them, people say oh, well that's what they mean by adaptive designs. The FDA will accept those ones, but not those ones.

So I think we need to be transparent about the selection process, but secondly, a careful choice across a range will send the right messages.

I guess if it was me, I'd go for -- if I was in the regulatory side, start off with kind of the easiest one you can, because just how to do the conversation, who comes to those meetings is going to take time, and then grudgingly get more complex.

And the thing I would really like to see done, but it may be too complex for a pilot, is actually that the work is put where adaptive studies are most needed, at chief of the public health, and to me you -- at last one area is in watching infectious diseases, where if you have a pandemic, and you want to be learning really fast, so that in real-time you get back. And I've been involved, and Berry Consultants have been involved in European package which is trying to get to that.

And to me, the public debating getting those studies setup so that they are ready to go, and so that you've got public buy in and you've done all the thinking would be well worth the investment. And I wouldn’t start there for the first quarter, but if you haven’t got one at all, that kind of thing, I'd be quite disappointed.

And then my final point is please make sure

that patients -- the patient voice, public voice, are involved in this, partly to get it right, and partly because of trust. But I think that that -- getting through the pilot, making sure that you've got that embedded from day one will pay dividends in the long-run.

DR. PRICE: Thank you. And Dr. Price?

DR. PRICE: Thank you so much. Just a couple comments primarily on discussion points two and three. Some of the type of trial designs I think that would be helpful are -- and we're also maybe per some of the disclosures that John was talking about, inferentially seamless types of designs, and borrowing methods, designs that have some form of formal borrowing, basket platform type of designs, many of these things that we've discussed today would be spot on with the types of things that again, have been done, but maybe not routinely. And so, we want to have more interactive discussion on how to do that.

Some of the factors -- and this is through a variety of conversations I've had as well as with some of my internal colleagues, what would be important to

help encourage, or not to inhibit a clear and simple process for submission of the pilots. If we get to a pathway where having a long drawn-out conversation about disclosure and what would be disclosed, that would cause some concern I think.

I mentioned this before, but just to reiterate, it would be important I think to have clarity on the reason if it is rejected. I think it would be a natural kneejerk reaction to say okay, whoops.

That's maybe not allowed, but we know that it may actually be that this is now the third you know adult borrowing for -- or pediatric borrowing from adult data, so that's really why, but clarify on why it was rejected and a pathway for continuing to allow that design to move forward, because it's important. And I think clarity on when and how the FDA will be communicating learnings, as well as routinely -- and updating that knowledge.

So this is going to be an adaptive thing, itself. So learnings on -- we talked about -- I liked the points about the time. There may be a window. You

may learn over time what timing seems to work well, and so having knowledge of that will be important.

And it would also be great to know if there are certain types of designs that you had hoped to see but you're not seeing, and communicating that I think would help sponsors to say okay, let's now look at where in our portfolio we might have an opportunity to come forward. So you might get some experiences that way. Thank you.

DR. PRICE: Dr. Emerson?

DR. EMERSON: I guess one aspect that I would want to see, were I doing this sort of thing -- I'm not. But is the process that you go through of what the initial submission was, and then what additional things FDA wanted to see for this documentation. What questions were not addressed in the initial submission? What things had to be addressed and that sort of thing? It's this process.

The word that's sort of missing from discussion point one to my mind is documentation. You're talking about the design. You're talking about the analysis, but what's the documentation of the

design and analysis that you'd wanted to see so that it was understood?

You said you needed to have the details, but it seems to me that this is a process of trying to gather information about how to in general present adaptive designs so that the FDA can understand them, or regulatory agencies in general, and what the sponsor would want to go with.

I guess I don’t -- you know from the start in a lot of this thing see that the clinical trial that's submitted ever has to be conducted really, so that there is a concept of if a sponsor doesn’t want to divulge too much of what they're actually doing, but wants to participate in this program.

It can be, you know, under not quite false situation, but some aspect there just to gain that information because that's the process that I would think that we're trying to identify the most.

DR. PRICE: So some of that, Dr. Emerson, could become a part of the selection eligibility criteria in terms of realistically planning to move forward with the study. So I'll move onto Dr. Zhong.

DR. ZHONG: I would like to respond to the comment that Dr. Emerson made. It kind of relates to question number one, this documentation, yes? For clinical trial design, it could in an initial submission, and definitely we -- there's some elements, like -- just like what we need to do, when we submit and initial hypothesis or synopsis.

It's not just a very high-level concept. We have to understand the patient population endpoint, a subpopulation group, and post selection rationale criteria, so on, so forth, like those high level things. And then more details on the statistical element design, elements of the trial, the trial design elements.

So that's why you select this kind of trial, and what's the rationale behind use of that, and for the trial endpoint. And those should be described in the documentation, right?

And then for decision criteria, and we will need to have some decision criteria, also described there. There's -- but not talk about simulation. It's a hard decision too, based on simulation to talk about,

all right?

And this particular method to analyze data and also make it through a decision, those things should be described in the documentation, and also the level of substantial evidence. So here we have this and for the other, we'll talk about another choice on disease in small population, right?

The evidence, it may or may not be the so-called the 5 percent offer level anymore. So we all have to open mind and think about how to assess the level of evidence based on the risk-benefit ratio. I listened yesterday in the rare disease workshop, and they discussed how to treat a patient who had a rare disease and no options. It's also a risk.

So we assess -- like I mean what evidence we need, then please look at the risk-benefit there. Yeah, I'm not talking about lowering the ball, lowering the criteria to get the job approval, all right? But we just have to assess the level evidence, it's not the five percent rule anymore.

And ideally, I mean to move the science forward and the documentation also contain like a

simulation plan and the evaluation of the important operating characteristics. So those things, I believe that we should have in the document.

So, but to the point number two, and here we think about the -- so maybe not point number two. So like in point number three, what factors might inhibit or encourage the submission for the program?

I feel like listening from Dr. Chan, we haven't disclosed certain information, but cannot like hide all the information. The agency can understand that. But I would like echo my colleague's point, like if we disclose too much, that could be a factor to kind of inhibit the submission through the pilot program, and would not allow us to achieve the kind of promoted innovation to the agency.

And also I mean the buy-in from the agency and those who are review divisions. And I think the horrible things that we spend months discuss this in a way to innovate design, but when we go to the phase meeting stage, or even latest stage, or else that review divisions said no, we don't accept this design.

But this could be huge factor there, that can

inhibit the submission there. So some kind of pre-specified enrollment in other division, review divisions into this review process, and could really help encourage the submission there. And also the guide -- career guidance is -- we're not asking for guidance for everything, but certain guidance still needed, even the guidance for the analysts require for this submission. It's needed.

I mean, I go back to simulation section and I fully understood that -- I mean there's no criteria for acceptance there. But I think that would be good that the agency could give some guidance on what a simulation plan should contain at high-level, right? What operating characteristics the sponsors should include in the simulation plan? So some guidance there would be helpful for sponsors. I will stop here.

DR. PRICE: So thank you. And Dr. Emerson, just to go back to your question -- I did want to make you aware that during the -- as we've been discussing the pilot program, one thing that we are considering is an active IND, so that also would guarantee that a company does intend to move forward with the design.

DR. EMERSON: But my point is if that's a barrier for disclosure, is that serving your goal? If the goal is to learn how to specify adaptive designs in a way that we can move forward and learn something that both.

It seems to be both industry and the FDA needs to devote resources to solving that problem and you're never going to do it if everything has to be you know proprietary information, and what aspect of this process, of submitting, thing like this.

Where do you even envision that the results of the clinical trials will enter into this? It really seems that this is an approval process and -- in what's going forward, and there's no mechanism whereby the results were there.

So is it really crucial that the clinical trial will be done, or is what's crucial is you know modeling after -- I would say industry would model it after something they're wanting to do, and in that process you present.

Now people sitting around this table have experience, coming up with documentation of adaptive

designs and the like. But what we want to do is get that out into the public domain and remove all of those barriers.

So, so far I haven't heard anything that says doing the clinical trial is crucial and we'll provide the information at the end.

DR. LAVANGE: So this is all hypothetical because the program hasn't been announced, but the discussion during the PDUFA negotiations, and Gracie correct me if I'm wrong, really had to do with -- okay.

The way it evolved was there was some discussion on the FDA side that sponsors were withholding protocol ideas, study design ideas, assuming that the FDA wouldn't take them. And we also felt a little strapped that we couldn't talk about some of the more innovative proposals for protocols we'd seen because things took a while.

And then, on the sponsor side, the feeling was that the more complicated designs took a lot of regulatory interaction because when you get into simulations you know just one protocol review -- I mean it takes a bit of back-and-forth, you know?

This is not what the FDA wants, or we need more simulations, or we need this, or we need that. And so there was a desire for more regulatory interaction. And so, we came up with the pilot idea mutually to offer increased interactions on a pilot scale, because resources at FDA are limited and reviewers are often fairly strapped for time.

So could we give -- could we pilot this? Could we try to give more interactions on a complicated design? And in exchange we can at least talk about the design so that people are aware that we are accepting or engaging about the protocol?

So we -- this is just the first step. I mean I agree that what you want is good, but we're sort of taking the first piece of this. So this would be a protocol review. This is not an NDA submission. The trial is not finished. This is a protocol review, and we're trying to enhance that process. Is that a fair statement?

DR. EMERSON: But it is the abstraction of this whole process that one, what we've heard is saying gee, how do you parameterize these innovative complex

designs? How do you parameterize the simulations that you do? What are the areas of uncertainty that the regulators have, and how do we get that back?

And it seems to me that that's what's crucial in this pilot program, is to say that that communication, the back-and-forth that you have, that's where most the education's going to come from.

I mean the question is, is what did we not know to supply to the FDA, and what after the FDA asked a question that they suddenly decided, yeah that was a stupid question. We didn't really need to now that. You know that information is crucial for being too (indiscernible).

Now I'm making this up, but you know the -- you know there. You have an SGR. Are you an SGE or is it -- Frank are you an SGE or are you a --

DR. SCOTT: I'm SGE and Maria.

DR. EMERSON: Okay. So but it's the idea of saying can you take in people who are some -- you know neither academic nor FDA, but trying to abstract it and then hit upon what that abstraction of this process is that's acceptable for release to both the sponsor and

the FDA to -- you know where you have some people who are dedicated to trying to describe a common framework on what that educational system is.

Otherwise, what this seems to me is this is you're encouraging people to submit things that will be in a silo and nobody'll ever find out about.

DR. LAVANGE: So the purpose of the public disclosure was to not do that. But that doesn't mean the pilot's going to answer everything. And so, I think we probably thought or envision that in this process you would have more public discussion at professional meetings about certain designs.

You would -- I mean I have sat in meetings where sponsors have put up on the screen all these different types of adaptive designs that some of which are not complicated at all and said this is what the FDA doesn't like because they call them less well-understood in their 2010 guidance, and I wanted to say no, no. You know we didn't mean that.

So you'll have more open discussion of designs at public meetings and so forth. You will have -- I don't know white paper, something will come out of this

about how to submit simulations.

I mean, I think there will be lots of things that will happen as a result of this. I think we're just trying to move the needle initially in this way, but I -- everything you say is a good thing to happen. I'm -- I don't think this pilot can make all that happen, but maybe it'll be the start is just my opinion.

DR. PRICE: You've stimulated discussion and Dr. Lewis wants to respond as well.

DR. LEWIS: Yeah. And just -- I just want to apologize if I seemed to make it appear as if I was saying everything in the program had to be fully disclosed.

In my mind, the -- one of the strengths of the proposal is the number which looked like two per quarter, which by my math is eight a year. And that allows for some diversity in the types of sponsors, the degree with which it's definitely gonna be run, the degree with which you can negotiate public disclosure of the communication and the iterative process, et cetera, et cetera.

So in my mind, I think the panel has identified a whole bunch of positive attributes of types of projects that might be accepted into the program, and the goal is to have as many of those attributes represented over the population of projects as opposed to presented in every one.

MS. KRAUS: So a few points for consideration on disclosure. Companies like government agencies can tend to be fragmented in terms of the statistical staff, the clinical staff, the regulatory staff, the legal staff.

When you're considering about applying for the program, you prob want to have had those conversations with your regulatory staff and legal staff to make sure that everybody's on the same page about how they feel about disclosure so that we don’t get a design that we really like, and then people go back to their regulatory folks and legal folks and say you can basically disclose nothing about this.

The other point that was raised about disclosing or requiring to disclose too much as being an inhibiting factor, I would encourage you to think

about what too much means under specific circumstance. You can approach it from the perspective of well you're asking us to disclose too much, or this is too much, or you could approach it from the perspective of let me assume that I'm going to disclose almost everything about my trial.

What do I want to hold back and why do I want to hold that back? What's the interest that I'm trying to predict in that process? And that may help you realize that maybe some things that you think is too much information really isn't too much information.

DR. PRICE: Thank you. Dr. Marchenko?

DR. MARCHENKO: I just wanted to know that of course level of confidentiality will always vary from sponsor to sponsor, and you will need to decide if you select this program because sponsor doesn’t want to give you enough information to incorporate confidentiality.

But what I do think would be important is to make sure that the least therapeutic area is announced because right now the perception is that those complex adaptive designs can be used only in oncology, or rare

diseases. And I think this is what the -- we would like to break too, because there many other therapeutic areas where those type of designs are applicable.

At the same time, I do want to agree with everybody's here that you are not trying to lower the standards. It doesn't mean that right now in every therapeutic area we are going to use external control because it's not going to be the case.

I think what Scott said -- I think it was important with regard how you're going to present the argument. Like from my perspective, it would be important that you would put initial design at least comments provided by agency, comments address the -- by the sponsor or by society writ large, so that it would be step wise process, not just here's the final design and that's what they would consider to be a good design. We need to understand why.

And what was discussed -- and then it's probably a question more for you. When I first read this program, I thought we were talking about confirmatory trials, here. But the more we discuss it, I realize that you're practically encouraged to be in

any kind of a design, even the interesting phase II design, which sponsor is willing to share. Is that the case?

DR. PRICE: So I think in its initial inception it was envisioned it would be more towards the confirmatory space.

We are open to innovation, but I will say during the initial inception, it was considered in the confirmatory space. Dr. Berry?

DR. BERRY: So Olga sort of asked my question, but I'll ask it again. Are you looking to show this scenario where you receive say an adaptive design report, simulation code, the results, and people get to see your reaction to that? Or are you trying to go a step backwards and have an interaction?

And how do you create that design? Did you simulate multiple things? Did you compare them? Did you approach that? Do you want to get into that building of an adaptive trial and get eyes on that, or just this point of what happens when somebody submits an adaptive design and get eyes on that?

DR. JOHNSON: So in many ways it's a little

bit of both because realistically you can't really do a good job of the first one without an understanding of the second. So I don't -- I would not expect -- it's a pilot, I should also, you know -- always the emphasis that part of this is that there will be a learning experience and a 360 structure here. It's going to happen.

But it's -- you can't -- I would hope that what would happen here is that something doesn't come in. And it's just a yes or no stamp, right? There should be iteration and discussion and that happens across every single type of design that we see, typically.

So I would expect it to go that way, but many times you have to understand like what were you thinking? Did you think about this? So I do expect there will be some back-and-forth, but again it's a pilot. That may change, and it'll really depend also on what comes in, but I would say you can't do the first without the second.

DR. BERRY: Yeah. It would be great to see some of that. And just a warning, I tend to be

somebody who is less concerned about operational bias and outside information.

At some point you may need to not disclose anything about a phase III trial that could give people information about how the trial's going, based on knowing the gory details of the actual simulated trial.

DR. PRICE: Dr. Meurer?

DR. MEURER: (Off mic.)

DR. PRICE: That's okay. You haven't said a lot so -- you haven't spoke a lot so I wanted to give you an opportunity.

DR. MEURER: Okay. Very good. So I guess I would like to I guess put in a plug for academic sponsor investigators, as a potential population for this program since I've collected a few IDs and IMDs as -- in that role.

I think that that also in terms of having the goals, of having a very open and transparent protocol. It may be easier in the academic space, particularly if it's repurposed existing approved drugs and new disease.

QUESTION: Do they have to pay the PDUFA fee?

DR. PRICE: No.

DR. MEURER: Not for an academic sponsor investigator.

DR. PRICE: Not for an IMD.

QUESTION: No, but to be a part of this program?

MS. PRICE no.

DR. MEURER: So I guess I'm interested in thoughts on I guess that population, for this. I mean obviously, you don’t even know how much excitement you're going to have about this, but that may be a population that could benefit from this process and would have an incentive since it would be government funded in the end to be doing very transparent work that could be seen by everybody.

DR. PRICE: Thank you. Dr. Mehta?

DR. MEHTA: Thank you. I think it's going to be very helpful when you actually make the announcement if you have some concrete examples of what types of designs will qualify for this program.

As this is going to be very complex designs, like this umbrella or basket type, or the simpler types

of adaptive trials, just to classify as not well-understood, such as just sample size re-estimation, or population enrichment, or some -- drugs, dose selection.

Even for these simpler types of adaptive trials, there are companies wherein the regulatory departments concerned that they will -- that they won't be acceptable to the FDA.

So I would strongly recommend that you actually list some examples in your announcement of the types of designs that could qualify for this program.

DR. PRICE: Dr. Harrell, followed by Dr. Zhong.

DR. HARRELL: One partial way to address some of these issues, especially Scott Emerson's issue, is to have a dry run with a hypothetical development, and hypothetical, clinical trial, that you could -- or a group could make up that sort of typifies a lot of the issues that you're going to see in real trials.

And then you have some sort of national webinar with lots of participation where you actually, in a crowd, start designing the adaptive trial. And

then you start gathering the general issues that come up, that Scott was asking about, abstracting those. It'll help the abstract.

DR. ZHONG: So I'll respond to Scott and Stefanie a little bit on the disclosure. And the post pharma and IO was in had members from the working group for that the applications, should be amendable to encourage popular learning. We're not going to try to hide from the public.

So I want to make it a bit clear on what -- I mean the group do not feel comfortable to disclose. For instance, I mean it -- like the indication, like the group doesn’t feel comfortable with that, might be similar action more towards structure right?

Then, even sometimes in the sponsor name -- a lot of sponsor do not feel comfortable with -- more company, one compound per disease. You disclose the name of sponsor, it could be a disadvantage to the company as well.

And study group are an sample analysis that performs on that status. So those are the things that these sponsors did not want to disclose.

The most important thing on the study designs, and other companies actually are willing to use these study parameters. So I just want to make a little bit clear there.

We are not trying to hide anything there. I mean the -- most of the company would like to participate in the pilot program, of course benefit company itself, but at same time also promote public lending, and promote innovation in statistic and clinical environment, so --

DR. PRICE: Dr. Lieberman?

DR. LIEBERMAN: Thank you. So one thought about some of these -- how to disclose some of the information that everybody's asking for, so instead of talking at a conference about one specific design that was observed.

The other one could be like quarterly, or biannual updates of here are the types of therapeutic areas that the designs will peak there. Here are the types of sort of discussions that we had, and it wouldn’t be tied to any specific design, but sort of summarized the issues that came out, the responses that

we were getting, sort of -- and put together.

So then everybody here is like you all got mention that it's not just the rare disease and oncology, but it is different therapeutic areas you will see it as, but also you'll see the variety of sort of responses and interactions, but not tied to one specific molecule.

MS. KRAUS: So -- introduce myself again.

DR. PRICE: They'll be connecting.

MS. KRAUS: Oh, apparently the people on the adobe can't tell who's speaking, so this is Stefanie Kraus, Counsel at CDER.

I wanted to pose a question to the panelists, based on the disclosure discussions we've had, which is when we get to the point of the disclosure discussions, so presumably something has been said in the initial application, we like the design, and we see that there is at least some opportunity for discussion, how do you envision those discussions going?

What we had not envisioned was that this would be a negotiation among lawyers, but a negotiation among scientists, so that we have an agreement on what we

feel is necessary to learn from these designs, and move the science forward.

So I throw out for you how you envision those going. Yeah?

DR. LEE: Maybe first suggestion would be that the sponsor in their application sort of provides. And she said that I helped her.

What I basically did, I took a protocol that was adaptive protocol, and I stripped all the information about the molecule, the therapeutic areas, the type of the endpoints, just maybe coded that this was type in vent endpoint, but I didn’t say that it was survivals and others, and said this is the statistical section. This could probably be shared anywhere with anybody, and the sort of types of adaptations.

So sponsors could in their application start with that, and just put this. Here's how its statistical analysis, or the synopsis: We strip the things that we don’t want to disclose. Hear that -- how it look like, and that being the sort of point is: Is that enough? And then maybe a little bit scientific discussion. Could we add a little more?

DR. PRICE: Dr. Lee?

DR. LEE: Well I'd like to make a general comment first. That is we often know less than what we think we know, okay? And if I count on anything today, I think learning is kind of the keyword, and we try to learn more, try to be adaptive, try to improve on what we -- how we do things.

Going back to this -- therefore I think it's great to have this innovative design pilot program, okay? And to move things forward in terms of applying better and more efficient designs.

But one thing more clear is that: How does this program works in terms of the -- in the drug approval pathway or the process? Because if it's already faced design, then it's less controversy, right?

But we heard this that is initially when it's planned is for confirmatory study. And we wish to understand this, it's not a MDA, okay?

So we know that, for example, in oncology that we have this accelerated approval, you know, based on indication, but need to be confirmed later for the

advocacy. So I'm not quite clear how does this pilot program fit in, in terms of the drug approval process?

DR. PRICE: So I'm thinking this through. Any of my industry colleagues want to help? Raji's going to help.

DR. SRIDHARA: So you know this has nothing to do with the approval process. It's an IMD that we are talking, or maybe even a pre-IMD, and all we are talking is about the design.

We haven't seen the results yet. Where is the approval decision coming in there? I don't think based on a design we are going to say this can only give you accelerated approval, or this can only give you a full approval, or what have you.

So I think approval is a totally different issue. All we are talking is we have a good design to answer the question that you want to answer. And how are you doing it?

The -- that's what is being done. And the other thing I -- if I can take couple of comments -- can I make? Okay.

The first thing is: I think all of you must

realize that all the trials, all on clinicaltrial.gov. So once an IMD is there, it has to be reported there. And I think of what more are you going to include in this pilot that you're going to disclose? Maybe more details about your design aspect and about the simulations.

And simulations are -- you know without your drug product there, and therefore probably there should be a little bit more openness in putting this forward. You could think of you know either there was a talk about how do we put this out, or what have you?

I think it we're not putting the disease in there, it's very hard for anybody to understand because the endpoints are quite different and why you decide and design is okay in one disease versus it cannot be okay in another disease.

There are very many factors that go into it, and so not having the disease in the background may not help always. You can always come back and say I don't know. Okay, time to move to the endpoint, but what time to event is you're talking about? Is it getting rid of an infection? What time it takes, or versus a

survival? There are very many different studies, so I think we have to look into that.

Some parts are present at how we should -- how we can put this out about the protocol or what have you.

And I want to think, and you know give us your opinion if this should be done somewhat similar to biomarker qualification program that we have, where we do have a website where one of it is qualified is put in there, and should we have a website where a trial which goes through the pilot, and the accept the design of the study can be put in such a way. What do you think about it? Is it too much?

And again, you know this has been on what you have already in the clinicaltrials.gov, and the age ability portion, and a -- I'm thinking that probably it cannot be static because if you're already have a design that's already in the pilot program, and it's made public, then another one coming with a very similar or close to the same design, do you really want to consider that?

So these are some of the things. And I think

from a sponsor point of view, you are to say when is it appropriate to announce this to the world, or to say okay, we can share. We are comfortable sharing at this point. Maybe when you put it in the clinicaltrials.gov, or after you have enrolled X number of patients, you feel comfortable? Okay. Our trial is running now. We don't mind putting this out.

So some of those things probably you know as sponsors, you might want to think about. Thank you.

DR. PRICE: And I would just add if -- I had to think about your question. It's not exactly a -- I think an easy answer to the question, but in the -- just because a sponsor is maybe not in the pilot program, that does not preclude an innovative design. And where would that protocol, or where would that design fit in the approval process?

So it's similar considerations, and some of that really is up to the sponsor in terms of where are they in their development program.

So I think just looking at the time, and seeing that we have someone at the mic, we'll go ahead and move to our audience participation, and we'll take

the person at the first mic.

AUDIENCE Q&A

QUESTION: So as I listen to this discussion on -- by the way, I'm Angela Johnson with CTI Consulting, and also Texas Tech University.

As I listen to this discussion I think of the many different types of sponsors we work with, some being very small biotechnology firms and some being very large. Then you've also talked about academic groups and also patient advocacy groups that come in.

With these first pilot trials, is there an interesting in going towards the platform route and sort of a follow-up question to that is: If so, does that favorite in the favor of one type of sponsor, and maybe even discourage say small biotechs from taking on these adaptive approaches, saying this is just for larger programs. This isn't applicable for me.

How do we make this equitable where smaller programs are also being brought in and saying yes, FDA is going to facilitate and have conversations with you about innovative designs when maybe you have a smaller pipeline.

DR. JOHNSON: Actually I could talk about that, but I think we're very pressed on time right now, so it might be better if we just go ahead and listen to all the comments and move forward.

DR. JOHNSON: So thank you for your comments. We'll keep noting them down.

DR. PRICE: So we'll go to mic two.

QUESTION: Hi. This is Dr. Liu from Bowling Green, department of statistics. We achieve learning form both positive and active cases. Sometimes I think we learn more through the -- I understand there will be incentive for sponsor who participate in the program and who are actually selected.

I'm just wondering if there will be any incentive for the sponsors to also share the rejections? You know you will basically select eight cases, and maybe you will reject 40 cases.

DR. JOHNSON: So we'll take that into account. Thank you.

DR. PRICE: First mic? Step --

QUESTION: Okay. I had a couple of -- some questions. I actually think clinicians want more

detail than the statisticians do when -- is it on?

PANELIST: Yes.

QUESTION: Maybe I'm short. Okay. So Roger made a good comment. I think clinicians hone in on certain details that they think really has to be there, and the statisticians tend to see oh, it's oncology with time to event. That doesn’t mean you know the kind of cancer, whatever.

But I can see the pilot program playing around with different approval rates, retention rates, dropout rates as part of a discussion. Because sometimes it will work for that kind of cancer, and sometimes it won't. And therefore it doesn’t really get down to the particular company's actual solution, and overall survival is something we all understand.

So there could be situations in which it's not a narrow indication. In fact, I think as part of the power program, we might want something that's not too tight because we kind of hope it's going to stimulate drug development.

And so therefore, it ought to apply in more than one little, narrow indication. But a lot of

things depend on what comes in and when. That's it.

DR. PRICE: Thank you, and we'll go to mic two.

QUESTION: Hi, I'm Katie (ph). I'm one of the clinical leads here at the FDA and I'm working in CBER. And part of why I wanted to work here is because of the innovation. I think the science is really exciting, and I think this panel's been fantastic.

Three practical suggestions for the care and pleading for your medical officers as we go through this process. One, I wanted to echo a comment I heard this morning about being transparent about the why, stating up front what's the rationale. So if I understand why a sponsor wants a novel trial design, it's a lot easier for us to agree on what. So just being really candid about that upfront.

You know we anticipate the biggest threat to the success of our product is this. We want this attribute in our trial to try and protect. Well now we can have a conversation.

And the second thing is show don’t tell. So a couple times I've had interactions with sponsors where

they're telling me what they want to do. I think I want to agree with them, but I can't tell exactly what they want to do. And I'll say send me an example. Send me a case report form with fake data. Show me how you want to score it and I'll look at it and say yeah, that'll be fine.

So simulations are good, but in this setting where it's really novel, you know an example, a choose your own adventure. Okay, if we encounter this state of knowledge A, we anticipate our trial unfolding in this way.

If we encounter state of knowledge B, we anticipate our trial unfolding this way. It's just sort of qualitatively describe different outcomes, depending on the scenarios you might encounter I think would really help to have a good discussion.

And then third is the assumptions. So you know I don’t understand probability, but I don’t need to to understand how a randomized control trial works, because I've had enough practice with them, but I understand the assumptions that we tend to bump into on whether or not the clinical scenario is likely to be

robust to those assumptions.

So the packages have to include some discussion of the assumptions you're making and why you think they're going to be robust in the clinical scenario so that we can comment on whether or not this makes sense. Thanks.

DR. PRICE: And the next person at this second mic before we move to the first?

QUESTION: Hi. This is Bob Beckman from Georgetown. And I just had a simple question following from the question about how this connects to drug development. And the answer was given that this might all be in the pre-IMD, or IMD setting, but I was kind of I guess this is a two part question.

First of all, I thought and I wanted to confirm, that it could also be available for a novel confirmatory trials, and if that's the case then I wondered whether a company that was willing to present novel confirmatory trial, if they went through a long interaction as part of the pilot program, and arrived at a mutual acceptable, confirmatory design, could they -- could that result in something that was equivalent

to a SPA?

DR. PRICE: So the spot has a lot of elements. So I'm hesitant to say yes or no. We really hadn't considered that. But I will say one part of the SPA that would not be attractive in this setting, in my mind, would be the time line.

An SPA is a special protocol assessment that has a very short timeline, and we're talking about fairly complex issues that will require significant time.

But if you're talking about an element in terms of just having some form of agreement, one would hope that if we reach a mutually agreeable design, that it would move forward.

That being said, that does not necessarily mean that the drug will be approved. You know you could do this great design, and it turns out your drug just doesn't work.

QUESTION: Right. Well that's true for SPAs, so --

DR. PRICE: It is. So we'll go -- thank you. We'll go to the first mic.

QUESTION: Thank you. Ruthie Davie (ph) from Metadata Solutions. Knowing that you may not have time to respond today, I just want to plant a seed for a thought later. And that is whether the agency would consider sort of hypothetical project, and I'll give you an example to explain what I mean.

So imagine there is a therapeutic area where a sponsor typically works with single arm trials, and going forward they'd like to use some sort of external control for those trials.

Would you consider a project that looks at a randomized trial, creates a historical control there, and sort of validates the creation of that historical control so that it could later be used in future drug development programs?

So the actual project is kind of hypothetical based around that randomized control trial. It's probably completed and removes a lot of the problems with revealing information.

DR. PRICE: Thank you for giving us something else to consider. And our last public comment or question?

QUESTION: Thank you. This Connie Latinagara (ph) from Eli Lilly and Company, and I want to thank you for -- panel members for all the great discussion so far.

We've heard a lot of technical topics as well as operational topics, today. And my question is: How do we envision handling you know all of these topics? Is it something that FDA is wanting to tackle themselves, or is it something that could be collaborated in nature, since we do have you know existing working groups, as well as taskforce with industry and pharma?

And what I mean by working groups is you know the adaptive design working group that have you know various complications so far, as well as the Bayesian scientific working group who have you know past work, as well as ongoing work to tackle some of the topics that we discussed today?

MS. KRAUS: Along those lines, just to remind everybody that we have a docket that's open, that's in connection with this public meeting. So if there's any points or comments you still want to make, or have us

consider anything, that docket is open through April 20th, and you can find the docket information on the agenda.

CLOSING REMARKS

DR. PRICE: So thank you all. This has been very helpful, and a very insightful discussion. According to the agenda, I am to give concluding remarks. These will be extremely brief so we can all get out and get safely to our homes. And I wrote a couple things.

So we started the day with the Swiss Army knife, and in my thought process on the Swiss Army knife was one size does not fit all, which was a theme that we heard yesterday at the rare disease workshop as well.

So there may be objectives and questions of interest that may be best answered via a simple, yet appropriate design, and there may be study objectives and questions that will require a Swiss Army knife that may entail a more complex design and analysis.

These complex innovative designs may entail adaptations, and may require simulations. The designs

may be based in frequentist methodology, or Bayesian methodology. Regardless of the methodology, the discussion today suggests that pre-specification is important in terms of modifications and adaptations that one might make.

General broad themes that I heard were the need for transparency, clarity, education, collaboration, including the cross functional effort, and communication.

Dr. Lee stated, and I wrote it down because I really liked it: education, innovation, and implementation.

So the education will be needed across disciplines within the FDA as well as among external stakeholders. A step toward educating in our opinion will be the complex innovative design pilot program, which will allow continuing education and information sharing.

This discussion today is not the end, but a part of the process of moving drug development forward, for again, the benefit of the patients and the overall public health.

The pilot program is a part of moving forward. However this does not preclude innovative designs outside of the pilot.

And Dr. Price, who -- we are not related, but I like that last name -- Dr. Price said we can do this. We must do this, and we are ready to do this for patients and for public health.

So again, I'd like to thank you audience for your participation. I would like to thank our panelists who came again, from near and far. We have panelists from the west coast. We have panelists from Europe, so this has just been a great discussion.

You've only seen a couple of FDA representatives today, but this really has been an effort among many of my colleagues, as well as many offices within CDER and CBER.

I'd also like to thank Dr. LaVange for -- while she's at the FDA, kind of laying the framework for this and moving us forward -- and for Dr. Estelle Russek-Cohen, who did the same as director of CBER, of the office of biostatistics at CBER.

I hope I have not forgotten anyone. I usually

do not like to name people, because inevitably I am going to forget someone. So that being said, thank you, safe travels to everyone.

(Whereupon, the foregoing adjourned at 4:15 p.m.)

* * * * *

CERTIFICATE OF NOTARY PUBLIC

I, KeVon Congo, the officer before whom the foregoing proceeding was taken, do hereby certify that the proceedings were recorded by me and thereafter reduced to typewriting under my direction; that said proceedings are a true and accurate record to the best of my knowledge, skills, and ability; that I am neither counsel for, related to, nor employed by any of the parties to the action in which this was taken; and, further, that I am not a relative or employee of any counsel or attorney employed by the parties hereto, nor financially or otherwise interested in the outcome of this action.

KeVon Congo

Notary Public in and for the

State of Maryland

CERTIFICATE OF TRANSCRIBER

I, Darcy Rogers, do hereby certify that this transcript was prepared from audio to the best of my ability.

I am neither counsel for, related to, nor employed by any of the parties to this action, nor financially or otherwise interested in the outcome of this action.

4/2/2018 Darcy Rogers

_____ _____

4/6/2018 Edited by: Benjamin Graham

z