US FOOD & DRUG ADMINISTRATION

_____

PATIENT-FOCUSED DRUG DEVELOPMENT

Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making

_____

DATE: Friday, December 6, 2019

TIME: 9:00 a.m.

LOCATION: Food and Drug Administration (FDA)
White Oak Campus
10903 New Hampshire Ave
Bldg 31, the Great Room
Silver Spring, MD 20993-0002

REPORTED BY: Michael Farkas, Notary Public

JOB No.: 3207312

P A R T I C I P A N T S

MEGHANA CHALASANI
THERESA MULLIN
SCOTT KOMO
MARTIN HO
KEVIN WEINFURT
GIANNA MCMILLEN
LINDA NELSEN
KENDRA HILEMAN
HYLTON JOFFE
LARISSA LAPTEVA
FRASER BOCELL
EBONY DASHIELL-AJE
BILL BYROM
MICHELLE CAMPBELL
ANDREA CORAVOS
MATTHEW DIAMOND
MARK FRASIER
ABIGAIL LUO
ANDREW POTTER
DIANE STEPHENSON

MARC BOUTIN

STEPHEN JOEL COONS

KATARINA HALLING

TELBA IRONY

LAURA LEE JOHNSON

PANDU KULKARNI

MICHELLE TARVER

MARY JO SALERNO

LAURA LEE JOHNSON

P R O C E E D I N G S

MEGHANA CHALASANI: Okay, everyone. I think we're ready to get started. We have a full agenda today, so we do want to get started.

Good morning everyone joining us in person and on the webcast. Okay. This time for real we're getting started, everyone.

Good morning. My name is Meghana Chalasani from the Patient-Focused Drug Development Program staff in the Office of the Center Director in the Center for Drugs here at the FDA. I'd like to welcome everyone to our public workshop. This workshop is the fourth in a series we've been conducting as we work towards developing a methodological, patient-focused drug development guidance series.

Let me first start by saying, wow. We have a very full room here, and I know we still have folks trickling in. And we've expanded from our typical use of two sections of the great room to all three sections, and we still have a very full room here. And I'm happy to see so many patients and

patient advocates, academic researchers, practitioners, medical product developers, and other key stakeholders in the audience. And I understand that we have nearly 200 or maybe even more by this point in time joining us remotely from the web as well. And we have many more registered who I'm sure will be joining us, both from here in the U.S., but also internationally. Thank you all for being a part of this very important workshop.

The purpose of the methodological PFDD Guidance Series is to facilitate the advancement and use of systematic approaches to collect and use robust and meaningful patient and caregiver input that can better inform medical product development and regulatory decision-making.

We kicked off the public dialogue on this effort when we conducted the first Guidance 1 workshop in December of 2017 and then continued the conversation last October during the PFDD Guidances 2 and 3 workshop. And so now we are here today to discuss incorporating clinical outcomes assessment into endpoints for regulatory decision-making.

Throughout the day, we want to hear from you on the approaches and considerations proposed in the discussion document for this workshop. If you haven't had a chance to read the document, that's okay. We will have a presentation to go over the key topics and set the context for us all.

For today's agenda, we will have Dr. Theresa Mullin, the Associate Director for Strategic Initiatives here at FDA Center for Drugs, get us started in the morning with opening remarks. We will then have a presentation that provides an overview of the discussion document for this workshop from Dr. Scott Komo, followed by a series of panel discussions.

We have five panel discussions today. The first one will be focusing on General Considerations for Developing an Endpoint From Clinical Outcome assessment data, followed by a second session on using the estimand framework to design, conduct, and analyze data from a trial with a COA-based endpoint.

The third session today will focus on considerations when there is a heterogeneity in

disease symptoms and functional status between patients and within same patient over time.

The fourth session we'll really work on discussing a working example, putting all the pieces together. And then we'll end the day with a session highlighting the themes that emerge throughout the day as well as reflect on the entire PFDD Guidance Series.

During each panel session, the audience will have an opportunity to ask questions and provide their views. We have a full agenda today. And for us to keep the conversation flowing, I am going to ask all of you to please be succinct and cognizant of the time.

We also have a public docket that will be open until February 4th, 2020, to which the public may submit general or detailed comments or examples regarding specific aspects of the discussion documents or topics raised during the workshop.

In the interest of time, during the audience questions and answers, our moderators may need to jump in to provide additional comments or flesh out your comments further through the docket or

discuss with our colleagues during the breaks that we've provided.

With our large number of webcast attendees, we will not be able to take comments or questions from the webcast during the workshop live. However, we also encourage our web stakeholders joining us via webcast to submit comments through the public docket. And we will take back any of the comments that you're putting in through the webcast after the meeting to review as well.

Following the panel sessions today, we will provide time for open public comment. If you wish to sign up to speak during this period, please do so at the registration tables outside. We'll have the sign-up sheets available through lunch. Participation is on a first-come-first-serve basis.

And so with that, I would just like to close with a few brief housekeeping points. We have an hour-long lunchbreak at noon. We do recommend that you preorder lunch. If you have not had an opportunity to do so yet, we will make that service also available at the ten AM break that we have. And

then -- sorry, 10:30 AM I believe is our break. And then we have an afternoon break at two PM as well. The food and beverages are available, as I mentioned, at that kiosk. If you aren't able to preorder, you will have an opportunity to do so throughout the day as well during the breaks.

Bathrooms are down the hallway in the lobby and on the left. The Wi-Fi password is up here on the slide. If you have any issues connecting to the Wi-Fi, you can also reach out to our colleagues at the kiosk and the information desk out front, and they'll be able to help you.

And with that, I ask at this time that you please turn off or silence your cell phones. And I'd now like to invite Dr. Theresa Mullin to the podium to provide opening remarks. Thank you.

THERESA MULLIN: Good morning, everyone. To everyone in the room, thank you for joining us today, and also for those of you on the webcast. I'm going to keep it brief. I think Meghana mentioned managing your time at least three times. So I'm going to try to heed that and give you a very

quick intro and just a quick recap of the background for these guidance documents. And this is the fourth in a series of four. They all have rather long names, but this one is about incorporating the COAs into endpoints for regulatory decision-making and some other related things that need to be addressed to operationalize this.

And this is really based on five-plus years, just to remind us all of where we got to this series of guidelines, from listening to patients in patient-focused drug development meetings and other kinds of venues over the last several years, recognizing and doing this that patients are uniquely positioned to inform us about the clinical context of the treatments that they will be taking and the disease context. In those patient-focused meetings, we took a systematic approach to getting their patients' perspective on the burden of their disease, of available treatments, and what they would most value in new treatments. It was a very powerful, eye-opening experience for us. And at this point we've had 26 FDA-sponsored patient-focused drug development

meetings and externally-led meetings. We're now up to 30 of those meetings. And that's quite remarkable when you consider that we just began having patients began running these meetings in 2016. So we've learned a lot.

And what we keep getting reinforcement of is this message from -- these takeaways about not only patients with chronic and serious conditions, but patients with acute conditions or maybe even less-serious conditions are really experts in what it's like to live with those conditions. And in talking to us about their experiences, we really have come to realize that often their chief complaints are not being factored into the drug development programs explicitly and as much as would be desirable, and including the measures of benefit and even better measures of what risks and what burdens patients are experiencing.

They've indicated they like to be as active as possible considering that they do have lives to live, families to raise, jobs to hold down, et cetera, but they want to help advance the development

of treatments for their disease. And these meetings -- and typically what we would hear after the first few meetings where there's always very powerful, a lot is shared. Patients would come up and ask us, okay, we've been telling you things that we haven't even told our doctor; what are you going to do with this information? So what's next, FDA?

And we really took this very much to heart. We thought these meetings are great and they are a very powerful source of information. But what else? What can we do next? And so this series of guidance that's being developed is really to help stakeholders get beyond and build on that initial and maybe qualitative narrative so they can collect information that could be used as endpoints in studies.

And in this picture we show it really informing the COAs, Clinical Outcome Assessments being used to inform benefit. But certainly they also can inform risk and safety and the burdens associated with the treatment. And so this is really where we're headed. We want to not only have the benefit of that

qualitative information that we would get in a patient-focused type meeting or a patient listening session, but really have it going to the next step and being data, become data that can be used for decision-making.

And so this is the series of guidance that we committed to do in a very similar kind of framing of this series of guidance is put in 21st Century Cures, Section 3002, the first one on collecting comprehensive patient input on the burden of disease. Moving on to the next -- and that guidance was published in draft in 2018 and we're working on looking at a final version of that now and publishing that in the next year we anticipate.

The next one, guidance two, developing a holistic set of impacts of burden of and disease and what matters most to patients. And that guidance is out in draft now.

And the third one will be on developing good measures of identified sets of impacts that can be used in studies. So developing or selecting clinical outcome assessment measures that would suit

and be appropriate for a particular concept.

And finally, this is where we're focusing today, on incorporating those COAs into endpoints that are considered sufficiently robust -- not significantly -- sufficiently robust for regulatory decision-making. I hate to say it, but it's even correct in the commitment letter. It should have been sufficient, not significant.

And so with that, here's the PDUFA commitment that we're satisfying in working through in publishing this guidance. Rather lengthy words. I'm not going to repeat all the PDUFA commitment language, but essentially it's talking about really putting something out that -- this Guidance 3 and 4 together in this series we think could potentially replace our 2009 guidance on PROs. And this covers a broader set, covers COAs more generally. And it would also address ways to incorporate those endpoints in decision-making.

The statute also asks that we put in information about standards and technologies to collect and analyze this information for decision-

making. So that's going to be covered in this guidance as well.

And so with that, I'd like to turn it over to Scott Komo so we can get into more of the substance of today's meeting. Thank you.

SCOTT KOMO: Good morning. My name is Scott Komo from the Office of Biostatistics at Center for Drugs. I want to thank you all for attending the workshop. Your input is crucial in developing this guidance on incorporating clinical outcome assessments, sometimes known as COAs, into endpoints for regulatory decision-making. I will now give an overview of the Agency's approach to this guidance.

This guidance will cover the methodologies, standards, and technologies to collect and analyze COAs for the purpose of regulatory decision-making. Guidance 4 builds on the work of Guidance 3, where Guidance 3 looks at the considerations when developing, modifying, or selecting a COA into a fit-for-purpose -- to be fit-for-purpose. Guidance 4 assumes you have a fit-for-purpose COA and discusses the considerations,

construct, and meaningful endpoint using the COA data.

As part of the discussion, it is important to note that the COA is not the same as the endpoint. We have heard this misperception, and I think it's important to clear up this confusion. As an example, let's assume you have a daily symptom score that measures the symptom severity over the past 24 hours. This is not the endpoint in your study. Instead, an example of endpoint would be the average symptom severity score over the last week prior to the end of therapy.

The guidance is targeted at a broad audience that I will now discuss. This could include stakeholders involved in design, conduct, analysis, and review of clinical studies incorporating COAs. In order to be useful for the statistical data management and related audience, (indiscernible) has been included in the discussion document and anticipated to be included in the guidance.

The audience includes medical product sponsors, clinical research organizations, consultants, academic researchers, FDA reviewers, and

patient organizations. Other audience (indiscernible) guidance include organizations involved with the development of registries, natural history studies, and endpoint or COA development.

I will now provide a brief overview of the sections in the discussion document. The first section is the introduction. It includes the overview of the four guidances in the PFDD series. This information is similar to that just presented by Dr. Mullin. It also includes a summary of the document. The next section is a discussion of the estimand framework.

The aim of this framework is to better align the study design, endpoint, and analysis with the clinical study objectives to improve study planning and interpretation of analysis. I'll give a brief overview of this framework in a few minutes.

The next session includes a discussion of the methods to determine meaningful within-patient change (indiscernible) their termination of study results. And finally, the last section includes the discussion of additional considerations which includes

topics on other study design considerations tied to blinding and the use of non-randomized or non-concurrent controls using COA-based endpoints. Also the use of computerized adaptive testing in clinical trials, and finally, formatting and submission considerations.

The document also includes an appendix, which contains two examples. The first example is a case study of the estimand framework looking at a COA-based endpoint as part of a breast cancer clinical trial. This example will be presented in Session 2. The second example is a gene therapy treatment.

I will now give a brief explanation of how the discussion document is formatted. Because of the broad audience of this guidance, we have used the following format. For the sections just discussed in the previous slide, the document contains a section summary that is aimed at a broad audience, a technical summary that is aimed at a more technical audience, and the technical details following in the section.

These details are derived in part from the reviewer comments that have been sent to medical

product sponsors and patient organizations. Several questions related to the document will be presented in the next several slides. We encourage you to provide feedback to the docket.

The first question, do you find this formatting approach helpful in understanding the material? If yes, do you recommend the guidance use this formatting approach? If no, what other recommendations do you have?

I will now go through a brief introduction of each of today's topics. The focus of Session One are factors to consider when constructing a COA-based endpoint. Some important factors to consider are you need to ensure that each COA-based endpoint is stated as part of the specific clinical study objective. Want to ensure that the COAs are fit for purpose and sensitive to meaningful change, the disease subtypes, the disease type, the treatment objective. Is the clinical study duration adequate to support the COA research objectives and also whether the frequency and timing of the COA administration is appropriate given the patient population, clinical

study design and objectives, and the COA (indiscernible) properties.

Other factors to consider would be ensure that the plans for scoring are specified and consistent with those used during tool development, including the handling of missing data. Also ensure the plans for COA measurement after discontinuation from treatment are driven by the research questions. Also need to consider the effect of lack of blinding for COA-based endpoints in open label or single-blind trials. And for non-randomized or non-concurrent controls, considerations could include whether the versions of the COAs are consistent between the study test group and the external control and also whether the COA administration methods are consistent within the study test group and the external control.

Next question. What factors should be included? Do you have additional factors that should be included and why?

I would like to introduce the estimand framework that will be presented in greater detail in Session Two. The first thing to note is that estimand

is the quantity used to define the treatment effect in a clinical study. The framework aims to align the study design, endpoint, and analysis with the clinical study objective to improve study planning and interpretation of analyses.

Here is a listing of the framework attributes that we discuss today. Important questions to answer are who is a target population for the study, what is the endpoint, how will events precluding observation or affecting interpretation be accounted for in the analyses, and what is the population level summary?

The following important issues to note regarding these attributes just discussed. These attributes are present in every data analysis and the choices made strongly impact the interpretation of the analysis, power, and data collected.

The document also discusses considerations and endpoint construction when there is heterogeneity in symptoms and/or functional status. This topic will be discussed further during Session Three. The heterogeneity could occur between

patients. Examples are having various disease subtypes where patients have differing symptoms and/or functional status, having a wide range in the rates of disease progression, a wide range in the baseline disease symptom severity, or wide age range of patients, which is especially an issue in pediatric studies. Or heterogeneity could occur within the same patient or time. Examples of this type of heterogeneity could be a disease with a waxing and waning nature or you have changes in functional status as children age. For example, as children get older, their walking ability normally increases, and this impacts the interpretation of the treatment effect on input like a six-minute walk test.

The following topics were included in the discussion document but were not directly discussed today. This is meaningful within patient change. This topic was previously discussed at the October 2018 PFDD workshop, so will not be discussed today.

Also the use of computerized adaptive testing in clinical trials and formatting and

submission considerations. If you have comments on the discussion document related to these topics, please submit it to the docket.

Please submit your docket comments by 11:59 February 4th of 2020. The topics could include but not limited to the content, the level of technical detail, formatting, examples for online materials, responses to the questions in the document. Also responses to the questions in this presentation. It is important to note that this is the last workshop in the PFDD series, so please include any additional comments for the guidance services. Note, if your comments are directed at a guidance other than Guidance 4, please clearly indicate which guidance your comments address. Please send in your comments. This slide provides the details and will be shown throughout the day during the breaks.

Thank you for your attention as well as your participation in this workshop.

MARTIN HO: Good morning, everyone. My name is Martin Ho. I am from the Office of Biostatistics and Epidemiology. Today we are going to

have several sessions. In the first session we will be covering high-level concepts. And in the subsequent sessions we will be having deep dive in the individual components of the discussion document.

So in the next sessions we are going to cover how to explore and discuss factors that need to be considered when developing a COA -- okay -- when developing a COA-based endpoint.

There are two points that I want to repeat, since we have to repeat the same point three times so that you can form memory (indiscernible). The first one is that endpoint is not the COA and vice versa. So I hope that over time you will recognize the importance of the distinction between the two. And the second point is that our discussion, we are focusing on the clinical trials or clinical studies that are being done, designed, and conducted for regulatory consideration. That is also a very important focus of the scope of the discussion.

So with no further ado, I would like to invite all of the panelists coming onto the -- taking a seat. And while they are taking their seats, I am

going to introduce them. Very happy to have a very good mix of expertise and perspectives. We have representatives from patient advocacy groups and also academics and sponsors as well as the FDA. And in terms of expertise, we have psychometrists, we have statistician, and we have clinicians.

Fraser Bocell is a psychometrician. He is coming from the Office of Strategic Partnership and Technology Innovation from Center for Devices.

Kendra Hileman, she is the Vice President, head of the Clinical Research and Development from Alcon.

Hylton Joffe, he is from the Office of New Drugs, CDER.

Larissa Lapteva, she is from the Office of Tissues and Advanced Therapies, Center for Biologics.

Gianna "Gigi" McMillen, she is a patient advocate and program administrator. She is from Bioethics Institute at Loyola Marymount University.

Linda Nelsen, she is a senior director

and head of Patient-Centered Outcomes at GlaxoSmithKline.

And last but not least, Kevin Weinfurt. He is Professor and Vice Chair for Research, Department of Population Health Science from Duke University School of Medicine.

Thank you for taking part in this panel.

So the first questions that I would like to ask of you is should the future guidance provide any additional details on the currently-proposed factors or considerations? First I would like to call on Kevin.

KEVIN WEINFURT: Hey. Good morning and thank you. And thanks very much for the opportunity to review this. And thanks to the FDA colleagues for all of your hard work. It was really a very comprehensive and helpful document I thought and will be of great use.

There were a couple of high-level issues that I noted where there might be opportunity for a little more detail or clarification. One was --

and I know we'll have a whole third panel on this. But the heterogeneous symptom and functioning presentation section was of course of most interest to me. And it was mostly things that will not work too well. And so I was excited to see more details about suggestions for things that might work well.

And the other, it was so helpful to have the first example, walking through the estimands and taking the time with a good example. And the second example was a very interesting example, and I wondered if we could get a little bit more structure about what aspects of the future guidance are being exemplified by different parts of that example. The first example was structured so beautifully and walked you through the estimand considerations. And so some similar structuring with the second one would be great.

But I thought that the considerations that were reviewed were relatively comprehensive and clear.

MARTIN HO: Thank you, Kevin. Next, Gigi, who represents the patient advocates.

GIANNA MCMILLEN: Thank you for having me here today. I thought this document was a clear presentation of many layers of information and I appreciate the logical sequence of the sections and the definitions of the key terms. It's still a bit unwieldy, and I am considering that if this document will be put in the hands of a patient or an advocate. A welcome added detail to the formatting of the guidance would be to indicate where patient input has been taken into account at each step of the process.

Now, I know there's a blanket statement at the beginning. But one idea would be to have a code or a symbol or a color cue that indicates the points in the described sequences when patient voice has already been incorporated into this plan and also where it will be needed in the future.

Even if this document does not actually enjoy wide patient or advocate perusal, I think that such indications of where the patient voice has been incorporated would add meaning to the descriptions of these unfolding processes.

MARTIN HO: Thank you, Gigi. Linda.

LINDA NELSEN: Thank you. I echo the previous speakers in finding this a very valuable initial draft of the approach to thinking about endpoints for COAs. And I especially value the emphasis differentiating the COA measure from the COA endpoint. It's a box that I jump on multiple times a day in my daily work. And the value of going from the concept to the COA measure, which is the way you're collecting that information from the patients to deriving meaningful endpoints.

And so I think one area that we can further clarify is really thinking through as part of the protocol design and development of endpoints is the nuances of the disease experience, the treatment benefit, and timing of the treatment benefit and that a single COA can create multiple, really valuable, insightful endpoints to help patients truly understand their treatment experience.

Scott mentioned a symptom diary where you get a 24-hour recall of symptoms. Well, you can use that single COA measure to derive so many different endpoints depending on the treatment

population. A mild population, you might look at symptom-free days. A severe population, you might want to look at change in severity. If you have it over the length of the trial, you might want to look at time to onset or durability of treatment effect. And so I think additional examples and emphasis on how to think through endpoints across all of those factors would be valuable in here.

I think that's all I had to say for here.

MARTIN HO: Thank you. Kendra, please.

KENDRA HILEMAN: Thank you for this opportunity to review the document and provide comments.

Especially because I come from a medical device perspective, most of the document kind of refers to drug, although the footnote has device included within it. So just maybe a bit more emphasis or inclusion in device in component of it.

And in particular there is a reference to real world evidence in one of the sections on non-randomization. But I see us having a lot more trend

toward studies that use real world evidence and endpoints based on that real world evidence, and maybe a little bit more inclusion of that information into the document would be helpful.

MARTIN HO: Thank you very much. Next is going to be the three colleagues from the FDA. First, Hylton, please.

HYLTON JOFFE: Good morning, everybody. So I agree with much of what's been said already. I thought this was a very comprehensive document. Actually, many of the concepts in this discussion document actually apply to endpoints in general when we think about designing trials. And we want to make sure the line up with the objectives of the trial, we want to minimize bias to the greatest extent possible, we want to have a pre-specified plan for analyzing the endpoint, controlling type 1 error and having a plan for missing data. So those concepts come through here. And then of course there are things that are unique to COA that are in this discussion document as well.

And I think at FD we think through a

lot of these things one-on-one with sponsors, so I'm quite excited to have a document that nicely in one place articulates some of these things that we are having one-off discussions with. And furthermore, when the guidance is published.

I would say that many of the things that come to incorporating the COA as an endpoint are not super-challenging. Of course there are some challenging issues. Kevin brought up a very good one, which is on the top of my list as well, which is when you have a lot of heterogeneity, for example, in the patient population.

So, for example, for seven years I've been a director of the Division of Bone, Reproductive, and Neurologic products. And we just published a draft guidance on interstitial cystitis bladder pain syndrome. Their patients have, you know, bladder pain or discomfort, and then they also have usually a lower urinary tract symptom. More frequent urination, waking up at night to urinate, urgency. But not all patients have all symptoms. So how do you best assess whether your drug is having an effect with an

appropriate endpoint? Because if your COA endpoint is measuring something that a lot of patients aren't experiencing or aren't experiencing much, you're not going to see much of an impact.

So I think the biggest challenge we see -- and I'll just take a moment to say this, I'm almost done -- is getting to this point where you're ready to pull the COA instrument into an endpoint. You know, how do you develop the fit-for-purpose instrument? I think that's where we have a lot of challenges. And getting those instruments as early as possible into trial so you can learn about those instruments early than Phase 3.

So in terms of additional details, I think the heterogeneity is a big one. The other thing I thought was -- on open label trials. I know you touched it on the example and there's a paragraph on blinding. But one dilemma we often face is in an open label trial is there any way that you can use a COA instrument, particularly if you're asking a patient about his or her symptoms. If it's unblinded, what's the framework around thinking about how that could

generate useful data.

Also another point worth making is making sure patients at baseline have severe enough symptoms so that you could see a reduction in those symptoms if the drug or device works. Because if you start off with very low levels of symptoms, there's not much opportunity for showing improvement.

And then my second-to-last point is also there's a section here about analyzing data. And I know Bob Temple has been on this for a while. We look at central tendencies, but I think it's also important to think about what the distribution of the treatment effect shows.

So, for example, we approve a drug for nocturia, which is waking up at night to urinate. And the mean treatment effects were pretty small, but we had endpoints there that looked at the patients who had zero episodes of nocturia, and that provided some additional useful context for gauging whether the treatment effects are meaningful or not.

And then lastly, again, putting in a plug to try and get COAs as endpoints tested earlier

in development. Like in Phase 2 it can help you consider guidelines for what's meaningful change. You can kind of see how the endpoint -- you know, how it performs in early trials before you then go ahead in Phase 3. So I'll stop there.

MARTIN HO: Thank you. Next, Larissa, please.

LARISSA LAPTEVA: Thank you. So I've read this document. And what I'd like to do first is I would like to commend the authors of the document for putting together a number of relatively complex and highly technical topics and going through this I thought very challenging task of presenting it to the stakeholders who may or may not have exposure to clinical research on a daily basis like some of us here. So I thought they've done a really good job with that. I am sure they will receive a lot of both positive and demanding feedback today. But overall, I thought it was a pretty good document. I would like to make a few comments and save one for later.

And so my first comment is about other elements that you would have to take into

consideration when successfully transforming a clinical outcome measure into a study endpoint. You know, we're in the business of drug development. And in no way anybody who has ever designed a trial would take only disease manifestations and only the COA measure, only those elements as the basis for how to transform a clinical outcome assessment into an endpoint. You will absolutely always think about what is the product that you're investigating, what is the treatment, and what would be the effect of that treatment.

And so one of the contextual elements in here is the effects of the investigational product. And of course the document is made applicable to a variety of medical products -- biologics, drugs, devices perhaps -- into different therapeutic area. But in some of the places throughout the document, I thought it may be helpful to mention that effects of the product would need to be taken into consideration.

For example, in the timing for assessments and the clinical trial duration and COA-based endpoint, when we're talking about things like

recall period or anticipated rate of change in the underlying construct to be measured, it is I think appropriate to mention the effects of the product when the (indiscernible) occurs, when the peak is expected, when the effect may plateau, and how long it would last. Because all of these things would be taken into consideration for applying a clinical outcome assessment measure to an endpoint measurement.

Another potential place is when we're talking about the choice of a clinical outcome assessment. Multi-domain responder index can be applied and connected with products when certain effects would be anticipated. And by the time when the estimand framework is considered, this is typically later-phase studies. And with these later-phase studies, you already have some anticipation of how the treatment might work. So for something like a multi-domain responder index, you would probably want it to be used in assessment of effects of a product that may have multiple effects. When it may influence say a uniformed or a very central pathway, like inflammatory response pathway for example. Or if it

is a substitution of a protein that will eventually downstream have effects on multiple organ systems. So that's one comment that I would make, is to ensure that the effects of the product are not overlooked, because they are a very important contextual element.

The second comment that I wanted to make is an expansion of what Gigi said earlier. We do talk a lot about how we could engage patients, how we could bring them to the table, how they are equal partners in product development. But in doing so, I think we should really be able to make space for this participation. And by having these discussions and publishing guidances and engaging patients, we are making space. But also when it comes down to this pretty technical aspect of product development, it would be also important to bring patients perhaps even at the design stage when they could be members of research teams and help out with some of these aspects. Because their input could be really very, very valuable.

Again, for example, in section where we talk about intercurrent events. And here we have

practice events and participant burden. And in the example at the end of the document, there is this talk about gene therapy, multi-luminance mobility testing, MLMT, and how to avoid the practice and learning effects. We had 12 different investigators in the study had 12 different configurations of the obstacle course assigned randomly to patients. And that's how they tried to avoid the practice and learning effects. So this is something where patients could really bring a very important perspective at the time when this clinical outcome assessment is taken and applied as an endpoint.

I'll tell you just very briefly a story from my own life. I used to work as a clinical investigator some time ago before my job at FDA. And I was a lead investigator on a study that was evaluating cognitive functioning in patients with autoimmune disease. And when you evaluate cognitive functioning, the outcome measure there is a battery of testing, of tests that have to be performed within the period of one-and-a-half hour. And before I started the study, I actually asked one of our clinical

research test administrators to give the test to me because I wanted to know what people have to go through when they take these tests and what the experience. You have the test and you have the instructions and you have various advice that's been developed for how to interpret the test. But until you actually have somebody who has gone through this, you may not see some of the very important aspects that would be useful in the interpretation at the end.

So I did the battery of tests. And after taking the test, I knew exactly how you can compensate one function for another and where in the test you can trade speed for accuracy and where you get tired and all the different stumbling points. And it was very, very useful for me at the end knowing what people had to go through and knowing their performance scores.

So, again, bringing patients as team members at the time of design I think would be helpful and useful. And if we could -- even though this is a technical document, but I completely agree with Gigi. Even if we have -- you know, if we can identify a few

places where patients' input can be of importance, I think we should do it in the document.

MARTIN HO: Thank you, Larissa.

LARISSA LAPTEVA: Thank you.

MARTIN HO: Next, Fraser, please. He is a psychometrician.

FRASER BOCELL: Good morning and thank you. I think I would agree with my colleague, Hylton, that it's quite a comprehensive document and covers a good bit of material. But at the same time, it does need to be accessible. And the goal is to make it accessible not only to professionals who do this for a living, but for other medical professionals who might be involved in medical product development. And so I think that even though it's comprehensive, it also provides just a starting point, a starting point for discussions to be had. Because especially in device trials and in mini trials, there's many challenges involved in including clinical outcomes assessment. And one of the key parts of this is this provides a starting point to engage with the agency, to engage with the centers and begin those discussions and

provide clarity in those discussions. And so I really think that this provides an excellent place and excellent information for us to move forward with. Thank you.

MARTIN HO: Thank you, Fraser. So we hear some common themes across the panelists, including patient -- laying down perhaps participation, how patient can participate in the process is important. Also that we would like to have more understanding or perhaps discussion on heterogeneity.

Last but not least is about the early and often interaction with the FDA when they are developing these COAs and then use it for clinical studies. And of course this is also important for us to get to understand how these effects are impacting patients in various aspect.

So the next question is what additional factors should be included in the guidance. Gigi, please.

GIANNA MCMILLEN: So I think the FDA is doing a good job with their intentions of including

patient input about all these different processes in drug development.

In the end, if you have this what I consider to be a pretty good draft of guidance, if you just drop it at the feet of a patient or an advocate, it's not very accessible or absorbable. My background is as a high school teacher. And so I get a little bit excited when I look at something like this. I think there's good information, there's many layers, it's compelling, the topic's important to the audience. How can you turn this into a meaningful teaching tool?

So I say okay, maybe -- the executive summaries that are in the document are in gray boxes. Those are excellent. But how about let's put them all together into a coversheet so that at one glance a normal human patient can just look at them and see what the context of the document is.

Also, I think patients need to have a statement about why this document is important and some guidance about where are the key points that they need to pay particular attention.

I think we're moving, I hope, forward to a place where meaningful instruction is not just paper-based. But I think a multimodal way of providing this information. For example, an interactive web-based document where you could click on terms and get a little interactive sort of description or video or something that explains each of the very important parts that this document covers. That would go a great distance towards having a fully-informed patient who, with this knowledge and with this understanding and with an appreciation of the context of their participation, that informed patient can more meaningfully participate as you gather this data.

So to keep my comments directed specially to the guidance document, I think there is good content here and it has great potential for being an excellent teaching tool. And that attention to that end will enhance the ability of patients to meaningfully participate.

MARTIN HO: Thank you, Gigi. I just want to echo what you have just said. In fact, I

think it's an excellent idea to be extended to all four guidances within the sequence. And that basically remind me some of the patient journey picture that we had presented by Dr. Theresa Mullin earlier. And I think by having perhaps a toolkit to demonstrate or perhaps to suggest points where patients can play a part from beginning to finish in the development process, I think it would be excellent. But it's my personal opinion only. But thank you.

Next, Kendra, please.

KENDRA HILEMAN: Great. So I guess the additional factor I would include is when you go through the document, it talks about different ways that you can evaluate endpoints or design endpoints as more of a you can do this or that or that. But what we find lately is that our endpoints are sometimes multiple factors within an endpoint. For example, you can have a statistical significance using a continuous variable analysis of difference in means. But perhaps the clinical relevance component of that endpoint might be based on a responder-type analysis. So what

we're finding is that a lot of our more current protocols have multiple elements you need to meet within an endpoint. And that complexity I'm not really seeing in the document right now.

MARTIN HO: Thank you. Next, Linda, please.

LINDA NELSEN: Yeah. I think one element that's missing that perhaps could have a little bit of attention is thinking about out are the way you're deriving and presenting the endpoints going to be things that are meaningful and important to patients. We do a tremendous amount of work to make sure that the COA measure itself is measuring important concepts to patients, but we make it very difficult for patients and providers often to interpret what that really means. And we think of COAs as a way of bringing the patient's voice into the clinical trial to quantify the important aspects of treatment benefit. And if we don't think through are these endpoints going to be interpretable, are there ways to display them, are there ways to explain them to patients so they can truly understand that

treatment benefit and use it to make good treatment decisions is incredibly important. And so I think we need to have a little bit of balance in the statistical rigor and intensity. And also are these endpoints going to be approachable and interpretable by patients.

MARTIN HO: Thank you. Next, Kevin, please.

KEVIN WEINFURT: I'm not sure what factors might be missing here, but I'll just share an experience I had. I actually had it this morning at breakfast. I was thinking about sexual functioning, which for me is one of my main measurement areas, so it wasn't that I was just thinking about that at breakfast.

And I really did -- I enjoyed this document. It's so comprehensive. But I was thinking we've got this measurement issue in sexual function where it's difficult to measure the functioning of someone over the last X period of time if they've not engaged in any activity. And so you get this interesting dependency. And past approaches have

tried to address that by assigning a zero score for functioning if you didn't have sex, and that has a whole lot of problems with it. Our (indiscernible) instrument takes a different approach and keeps having the activity and the functioning conditional upon having activities separate things.

And I thought, well, if I were to -- if this were the guidance in its current form and I were to go through and say how do I think the FDA would like me to think about taking the responses from our measure and creating an endpoint here from that. I wasn't sure that I could thread the needle all the way through there. I still have uncertainty about what would be an acceptable approach there.

And so I just offer that as one case example where it might be interesting to think of those and some other ones and go through and see do I come out of that with a better sense or less uncertainty about what I ought to do than I did before. So I would just offer that.

MARTIN HO: Yes. Kevin, I think it's a very good point. In the guidance documents we talk

about estimand. We talk about what estimand individual unit. But yes, it would be ideal to have perhaps a roadmap or a story-telling process or engagement process to think about these translation or development process.

Next, Larissa, please.

LARISSA LAPTEVA: Okay. So I would like to add and talk a little bit about the meaningful within-patient change. And I know that this is one of the topics that was listed as not supposed to be covered, but we were offered to provide comments about it.

MARTIN HO: Yes, yes.

LARISSA LAPTEVA: I thought this was a very important topic to bring up. And as I see in the documents, the section is written from the perspective of the mainstream clinical trialist who is doing comparative research and trying to assess within patient change and how it is related or not related or could be applicable to understanding and evaluating the differences between the treatment groups. And this is all great and in no way -- I want to make a

disclaimer. I am not a heretic; I am a scholar of the adequate and well-controlled and all the comparisons and the statistics that's behind it. But in certain situations you just can't have an appropriate or the best in the world comparison. And in those situations where you can't use a concurrent control or where -- and this may be something looking forward to -- where you have an effect of the product that is so compelling where within patient change, the meaningful within-patient change can become a very important metric. And so I think the within-patient change with any outcome assessment measure should be a necessary attribute of that outcome assessment measure.

And I saw the discussion here in the description. And there is methodology that's offered to be used. Very useful. In fact, we've seen it in practice. But I also saw that there wasn't a discussion about the variability in the disease and the variability for individual patient performance. And these are very important factors that would need to be taken into consideration, because the variability of the disease -- and I'm not talking

about just relapsing-remitting type of diseases. Even chronic diseases that are considered to be progressive may have certain periods in them where the disease stabilizes or can improve even -- we call them honeymoon periods on its own. Another aspect is that each patient may have better days or worse days for various reasons. And so here is this within-patient performance on a COA, right, variability.

So these are two factors that would need to be absolutely included I think in the meaningful within-patient change evaluation. And again, many of you have probably heard about (indiscernible), maybe you have seen publications about it. Trying to consider how we could potentially develop individual therapies for patients. And if you have a product that's tailored to an individual patient, then you may not have the comparator arm. There may be no study. And so in that case, the meaningful within-patient change may become one of the metrics that are extremely useful in evaluating how the product works in that patient.

MARTIN HO: Thank you, Larissa. As you

can tell, people from CBER were very mindful about how to evaluate and understand innovative products, as we are encountering quite a bit.

Now last but not least, Fraser will have the honor of wrapping up the question one and question two. He is the last panelist to share his thoughts. Thank you.

FRASER BOCELL: Thank you, Martin. I would like to kind of echo what Linda said about the interpretability of this and that when you're going into this, not only should the COA bet interpretable, but the analysis you do should present something that is then interpretable as well. So the endpoint needs to be interpretable when it goes into the labeling.

And then I also want to talk a little bit about some of the device-specific considerations that need to be made. Because this document is not the only consideration that should be made. There are device-specific guidances as well as device-specific standards that also touch on clinical outcome assessment as endpoints, the types of endpoints you need, and things that are necessary. And so it's

always important to look at this and use it as a guide in conjunction with other things that might be important within your product area, within your thoughts.

One example is interocular lenses that specify certain types of measures from patients that need to be done within their standards. And so these are things that you always have to keep in mind moving forward when developing your endpoints.

MARTIN HO: I apologize. In fact, I think Hylton will be the next person.

HYLTON JOFFE: I thought I was off the hook.

MARTIN HO: No, no, no. Not that easily.

HYLTON JOFFE: I've actually been wracking my brains trying to think what additional factors are missing. I interpret factors as the general categories that were laid out earlier. And I really couldn't think of any. I think there is further details that can be added into each of the factors that you have. I echo some of the other

comments about making sure the document is accessible to the primary audience. We had the slide of who the primary audience is for the documents. So just thinking through the perspective from each of those as to how to make the document most accessible.

And then the other is just adding in more details in some places. And this is a balance of how much details do we put in. But we talk about things like rescue therapies or intercurrent conditions effecting interpretation. But, you know, maybe adding color on how do you handle that exact -- maybe giving some examples of how you would handle someone who needs a rescue mediation that's going to impact the COA. So I don't think those are different factors per se. I'm thinking of them more as filling in some of the spots here and there.

MARTIN HO: Thank you. I have a follow-up question to Hylton. You and Larissa and Fraser gave excellent feedback about from your perspective what you would propose to pay attention to from the therapy, beginning to the end. Do you think in this guidance there would be a useful -- perhaps

description or discussions about when it comes to developing these COAs and then translating them or developing endpoint based on those COAs, what would be perhaps a typical engagement process for sponsors and for others when they want to elicit inputs from the FDA?

HYLTON JOFFE: That's a good question. So within CDER -- and I'm sure there's analogous processes in the other centers. The review divisions work very closely with our COA colleagues and also biostatistics. It's really a joint effort to evaluate these things, because no one group has all the expertise you need when you're bringing this all together.

We sometimes have what we call Type C meetings, which are meetings where we just focus on discussing these instruments, because there's so much detail in there. So we encourage sponsors to request those types of meetings if they really want to have a dedicated focus on their instrument.

We're trying to get the work on the instruments to happen earlier in development, as I was

mentioning before, and really try to get folks getting some of these endpoints into their earlier trials. Because if you have a phase 2 trial that uses one endpoint, and then your phase 3 trial is going to use a different one, that introduces some risk. Maybe it's not going to translate to the new instrument in phase 3 and you end up with a failed phase 3 trial for example.

In the OND divisions, our regulatory project managers are our point of contact. We encourage folks to reach out, request meetings if they feel that would be helpful. I think a lot of the discussions with endpoints going into phase 3 trials happens at the end of phase 2 meetings. But again, we are hoping we have information before that can really help inform how we think about the phase 3 trial development.

MARTIN HO: Thank you. It's very useful feedback. Yes, Larissa.

LARISSA LAPTEVA: So I think this guidance on its own, the fourth guidance in the series, is an important tool to help development of a

COA, because it gives the ability to begin with the end in mind. When somebody is interviewing patients and trying to figure out what are the important domains for a disease and then develop a clinical outcome measure and then validate it, you would need to know how it will be applied. And so this is how it will be applied. So this guidance on its own is an important feedback. So that's one.

The other is that, as Hylton mentioned, we have a number of pathways or ways to interact with sponsors, with those who develop products. And Hylton talked about how we could speak with individual sponsors within their individual development programs. This could be done very early, at even pre-IND consultation meetings. In CBER we have a former pre-pre-IND program. It's called Interact. Even there certain aspects of clinical development can be discussed for an individual product within an individual product development program. And then later in phase 2 of course, and then in phase 3.

So the other way to get assistance in development of a COA is to work with the agency

through the Drug Development Tools Qualification Program. I think both CDER and CBER now participate in this program. This is a process where people could develop a measure in what we call a precompetitive space where consortia or multiple stakeholders could get together and develop a measure for a particular context of use that could be applicable to more than one disease or more than one product.

There are other pathways, including Critical Path Institute and (indiscernible) for the innovative technologies in CBER. There is information about these different ways to interact online. If you have any questions about them, I'll be happy to answer after.

MARTIN HO: Thank you, Larissa. I also believe that there is a parallel opportunity to engage colleagues at Center for Devices as well. Given the available time for Q&A for the audiences, perhaps if you have any questions, please approach Fraser and our colleagues to discuss these potential (indiscernible).

So let me start with this audience Q&A. Please approach to the microphones in the middle of

the aisles if you have any questions. Okay, great. While we are waiting for someone to come up with some questions, I think one thing that really strikes me is that there are some discussions about people needs to flesh out a little bit about heterogeneity. And I want to see if Kevin can elaborate on his thoughts on this part.

KEVIN WEINFURT: Well, given we're just sort of commenting at a high level for that section -- and hopefully our later discussion will illuminate more aspects of this. But I think the section does a great job of addressing the challenges in trying to represent and analyze situations where you have heterogeneity in symptoms or functioning within or across people. But as a user of the guide, as someone who is actually trying to do the work, I've left with some options feel like they're going to be unfavorable, but I'm still left trying to figure out what to do. And it's really challenging. And so I don't expect that there was a magic solution hidden in the back room that they forgot to put in. And it's hard work to do. But to just highlight that that is

an area -- it's an issue in the guidance because it's an issue in the field. And so -- yeah.

MARTIN HO: Thank you, Kevin. I think Kevin is at the front line of research on this topic. So if he is still trying to wrap his brains around, then that means that a good solution is yet to be found. So therefore, I think it would be good for us to, as Kevin said, to mention these may be challenging topics. And perhaps in case-by-case basis we can engage with individual sponsors to figure out some fit-for-purpose solution for that.

And I think right now is our -- okay. So yes, Hylton, please.

HYLTON JOFFE: I was just going to follow up on this. I was wondering, maybe Kevin, do you have thoughts on using things like most-bothersome symptom or things like that as a way of trying to get around the heterogeneity? The document talks about personalized endpoints, and I actually wasn't quite sure what that meant exactly. I think that's worth clarifying. I wasn't sure if that meant each patient has his or her most bothersome symptom and that's

where we have a problem. I wasn't quite sure on that. But I don't know if Kevin or anyone on the panel has thoughts on most-bothersome symptom approaches.

KENDRA HILEMAN: Yeah. I thought that was probably the most confusing section for me, the personalized endpoint.

But one of the other things I wanted to mention about the complications with the different endpoints is oftentimes in the medical device area, we have a surgery involved in the study design, and that introduces a lot of different effects into what's happening with the patients. It oftentimes means that the clinical outcome assessment that's done before the surgery really has very little relevance to change from baseline for the patient afterwards. So sometimes we just look only at the postoperative outcomes between two treatment groups. And sometimes we do. It's often when that's the case when it's a surgical intervention, that's a good example of a study design where a control really isn't available. You know, we don't do sham surgeries very much in clinical literature. So those are some of the

complications that come with -- and what I meant by adding some device considerations into the document is it introduces a little bit more of the intermittent events and the heterogeneity across the patient across the study becomes more of an issue in some of those trials.

MARTIN HO: Yes. Kevin. Kevin first and then Gigi.

KEVIN WEINFURT: Yeah. And I know our third panel is going to get more into the weeds on this heterogeneity issue. So I don't want to do that. But at the high level it does seem like as that section gets developed and takes shape, there might be some opportunity to look back into the guidance that precedes this one where people are trying to figure out how to put together a measure. And this is an interesting area where some people are going to try to solve the problem through a scoring solution of the measure and some people are going to try to solve it by defining an endpoint in a certain way. And in my experience sometimes measurement folks like me are not even aware of some of the different options where you

could handle the complexity on the endpoint definition side as opposed to trying to come up with a scoring solution for it.

So this might be a place to earmark an opportunity to make some connections between the two guidances. So as I'm going through Guidance 3, trying to figure out how to make a measurement situation like this, I'm not assuming that I need to solve this problem in the scoring. I'm aware that I should think ahead through the estimand framework and be aware of these other options.

MARTIN HO: Yes, Gigi.

GIANNA MCMILLEN: I just want to point out that there's two different kinds of patient input here. You've got the patient who is newly exposed to this medical emergency and treatment. And so their evaluation of what is important or serious or bothersome could be different than a patient who is more experienced or further down the line in the path of their disease and treatment. So the more experienced patient has context and may realize that over the period of time what seems very bothersome in

the beginning is actually in context not such a big deal after all. So you have two different patient populations who are giving their personal input. Both inputs are valuable and valid, but they are different.

MARTIN HO: Thank you. So I see that there potentially is an audience who want to ask a question. Go ahead.

RACHEL LAWRENCE: Thank you. I'm Rachel Lawrence. I'm a statistician for Adelphi Values. So from a sponsor perspective. But I just wanted to have a listen to what the panel said today, and particularly perhaps Gigi's comments on what's highlighting in the guidance where the patient input is. I wondered if you could consider sort of also reflecting what the regulatory decision-making points are. So what is it -- which aspects of the guidance is really key, what are you needing at the COA endpoints and how does that feed into the overall risk-benefit decision-making part. It may be a question for the panelists to comment on.

MARTIN HO: I just want to add that, as you may have known, our agency has been working hard

on the benefit-risk part. And hopefully you will hear some good news on that. And in terms of the benefit-risk guidances and agency's thought on that, I think I would like to see if my colleagues, Hylton, Larissa, and Fraser have anything to say.

HYLTON JOFFE: Well, internally we are thinking through and having a lot of discussions on the benefit-risk framework and how to think of that as the endpoint in your drug development with that as the goalpost, maximize your chances for showing a positive benefit-risk assessment at the end of the day. If your drug has a serious risk, it's even more critical to really try and find the endpoints that are really going to maximize the benefit you're showing so that we can hopefully end up with a favorable benefit-risk. So I think it's very important to think about benefit-risk as where we're going and take that into account when you're thinking about the endpoints, whether it's COA endpoints or non-COA endpoints that are going to add to the benefit side of the equation.

MARTIN HO: Thank you. Larissa first and then we can go Bennett.

LARISSA LAPTEVA: Well, I only wanted to add that in response to the question of where in the guidance you could put some goalposts for potential patient input. I think most of the section in the intercurrent events can benefit from patient input. But patient input in drug development is much more complex and goes beyond this guidance. There are many other aspects where patients could be extremely important in giving their perspective of how to develop a particular product or how to design a particular study. So this is a much larger conversation I think.

MARTIN HO: Thank you, Larissa. Fraser, do you have anything to add? No, okay. Bennett.

BENNETT LEVITAN: Bennett Levitan from Janssen R&D. So I wanted to ask people for your thoughts on the comment that Kendra made earlier. Many times an endpoint is based on a continuous change in a COA measure. And that has a lot of fidelity, but it's also rather abstract. And often when we talk to physicians for getting to benefit-risk, they want to

see a measure of response, some threshold change. And that is often not measured with a statistical hypothesis test. And we're in an awkward position of doing benefit-risk with an endpoint that wasn't in the hypothesis chain.

So since you just brought up this benefit-risk point, the idea with the endpoint is not only to use it alone, but to use them collectively to render a benefit-risk decision. What are your thought about possibly suggesting moving towards the dichotomization of these endpoints so that they are more clinically meaningful for some of your physicians who are ultimately going to do a benefit-risk balance?

MARTIN HO: So before I ask for my panelists' comment on that, now I am wearing my statistician hat. And I have to say that in testing a hypothesis, continuous outcomes often are more powerful to detect the differences between groups than our responder analysis which are dichotomized. And in the process of dichotomization, we lost some information. So this is one part of -- one of the considerations of how to pick an endpoint. But with

that, I would like to see if my colleagues have additional comments.

LARISSA LAPTEVA: This is a question that comes up so frequently. It's almost like a chicken-and-egg argument sometimes with some endpoints. Continuous versus binary. And you've heard the perspective that Martin just shared, and that's one side of the argument. If you're using a continuous variable, you're absolutely going to lose on the accuracy of the comparison. If you're using a binary outcome, it may help a little bit mathematically and it may look like clinically it is pretty meaningful. But then how do you say separate somewhat responder from really high responder who has improved very significantly. So this is something that we encounter pretty frequently. And there is no universal answer to this question. We have used and advised in using and sponsors have used and drugs have been approved on both. It really depends on the disease. I know it's a disappointing answer, but it depends on the setting on the disease and how valid either continuous or binary variable would be in a

specific setting.

MARTIN HO: Thank you, Larissa. Hylton, do you want to have anything to say?

HYLTON JOFFE: Yeah. I think the continuous interpretations can be quite challenging. Things like cumulative distribution curves and things like that are not very accessible. Even folks like us who see them quite a bit have to always wrap our brain around them. It would be nice and ideal if you could use information from both to help -- you know, going back to the nocturia example for example, bringing in -- so this is when you urinate at night. You know, having no episodes or having at most one. Most folks think that having up to one episode of nocturia, some folks think that that's actually completely normal. So grounding a responder analysis in something that makes clinical sense, and also pre-specifying all this. Because when you start cutting the data in different ways, you can get different answers. So I think it is a challenge. And in the ideal world it would be nice if we could get information from both. But I think Larissa makes a good point that the

specifics of the development program, the disease, and so on and so forth may move it towards one or the other.

MARTIN HO: Yeah, thank you.

LARISSA LAPTEVA: The key -- sorry.

MARTIN HO: Larissa --

LARISSA LAPTEVA: Just one sentence to add to this, that consistency is important, if the two are consistent in their performance, responder versus continuous, then this is something that provides reassurance there. You know, measuring the same thing.

MARTIN HO: Thank you.

FRASER BOCELL: And I was just going to say that we do look at the totality of evidence. And you're also thinking about not only what you're submitting to get approval, but then what's going to go into your summary of safety and effectiveness, what's going to go into your patient brochure, what's going to go into your physician documentation. And so it is something where you can focus on what's going to be your primary endpoint, but then specify secondary

endpoints or other things that they're going to use for that further documentation. So it's really talking to us and seeing about how you want to approach that and what your goals are and what we can look at and do.

MARTIN HO: Thank you, Fraser. So we are just three minutes past. And last thing, our outstanding panelists. Okay, I didn't know that. Please.

CARRIE BARNHART: Hello, my name is Carrie Barnhart. I am a patient advocate. I used to be a teacher and I used to work in pharma. I used to be somebody, and now I'm just here on behalf of all rare disease patients.

And going off of what Ms. Larissa was saying about every patient is different, we have good days and bad days, with rare disease, often we are misdiagnosed, we have comorbidities, we have many, many, many diseases, especially with autoimmune disease. So oftentimes there is nobody driving that ship. So how are you -- I don't see in the document how patients like myself and other patients with rare

disease are going to be contacted if there's no specialist for that disease. There's not -- it doesn't fall under rheumatology, it doesn't fall under genetics. So I'm just wondering how to include patients that fit outside of the box.

MARTIN HO: Any comments from our panelists?

LINDA NELSEN: When we're doing certainly COA development, we are beginning to look at more innovative ways than recruiting through individual patients to bring patient voices in. We are looking at social media to have online discussions, we're looking at communicating with patients through sort of advertisements perhaps on Facebook and other ways where there might be online communities that we can tap into. And I think it's a really important way to make sure we have a much more diverse population when we discuss, make it easier for patients with limited mobility or limited stamina to participate in these discussions. And so I think formalizing and finding acceptable ways to use those in regulatory product support is important.

One of the biggest areas that we fall into when we do this online research is the FDA obviously want us to have confirmation that a patient has a certain diagnosis. And if we're interacting with someone only through an online interface, it's very complicated and it takes many steps to get that confirmation. And so that's a challenge for us. But we like these ways of interacting with patients outside of standard clinical trial or clinical sites. You may be more comfortable discussing your issue with us when you're in the comfort of your home. You may have more time to think about it. You may have more time to look around and realize a lot of concepts that when you're in a face-to-face interview won't necessarily come through. So I think those are ways to reach out to patient groups. The online communities, and patient advocacy groups are one way we can do it.

MARTIN HO: Thank you. Fraser first and then Larissa.

FRASER BOCELL: So at the Center for Devices, I am part of the Patient Science and

Engagement Team. And while I am more of the patient science side, we do have a group on our team that deals specifically with patient engagement and including the patient voice. And so that's something that -- that's an avenue, and we have several different avenues throughout that that provide the opportunity for patients to become involved and to include their voice in different aspects.

MARTIN HO: Larissa?

LARISSA LAPTEVA: So we collaborate with a number of patient advocacy organizations. And you may know that there is a National Organization for Rare Disorders, NORD. And NORD has a patient registry where patients themselves -- this is not about data entry done by physicians or investigators, where patients could put in their data. And there is a large network of connections with natural history registries, with those who may be looking in an investigational way or collecting data about very rare disorders. So that is I think a good network and a good avenue to tap into for something like an outside-the-box type of condition.

MARTIN HO: Thank you. So in the interest of time, I think let's wrap this up. And the gentleman over there, please feel free to contact our panelists during the break time.

So the next session will start at 10:45. Please come back. Thank you.

(Break)

MALLORIE FIERO: All right. So in the interest of time, because we're running a little bit late and I want to make sure that you all can get to lunch on time, because I also want to get to lunch on time.

So good morning, everyone. My name is Mallorie Fiero and I am in the Office of Biostatistics in the Center for Drug Evaluation and Research. And before I begin, I wanted to acknowledge that we originally had Allison Campbell as part of our panel today, but unfortunately she could not make it. So we have Kevin Weinfurt, who was just in Session 1, that stepped up for us today. So we appreciate him and his feedback.

So for Session 2 -- or I'll step back a

little bit. In Session 1, we kind of talked about broad considerations of this discussion document of the COA-based endpoint. I heard a lot of really good discussion there, so I hope that we can continue on with our discussion in this session.

So in Session 2, we begin with a fundamental issue that must be addressed for COAs, which is to ask ourselves, what is the question that the clinical study is designed to answer. If we are not in alignment on the right question, interpretation of COA results will be difficult. The estimand framework is based on the ICH E9(R1) addendum and aims to improve alignment of the research objective with the endpoint analysis. And this will help us to improve interpretation of results. Okay?

So now, I know a lot of you who are non-statisticians are thinking like, oh, she's talking about the estimand framework, this does not apply to me. But this is false. This applies to all statisticians and non-statisticians because this framework aims to improve dialogue between all disciplines who are involved in the objectives,

design, conduct, analysis, and interpretation. So this is also important for you, too.

So in this session, we will re-introduce the estimand framework and then we will apply it to a case study using physical function as an example in an advanced breast cancer trial setting. And in this case we aren't detailing any specific estimand that you should pick, but rather we're talking about the conversation that should take place, this multidisciplinary conversation that should take place early on.

So we have a great set of panelists for Session 2 today. I'm going to ask each of the panelists to please introduce yourselves by telling us who you are and where you're from. And I will start with Jessica Lee.

JESSICA LEE: Good morning. My name is Jessica Lee. I am in the Division of Gastroenterology and Inborn Errors Products in OND.

GREGORY LEVIN: My name is Greg Levin. I am in the Office of Biostatics, Center for Drug Evaluation and Research, FDA.

JOHN SCOTT: My name is John Scott. I'm in the Division of Biostatistics in the Office of Biostatistics and Epidemiology at the Center for Biologics, FDA.

DANIEL SERRANO: Daniel Serrano. I'm the Director of Psychometrics at Pharmerit International.

KEVIN WEINFURT: Kevin Weinfurt. I'm with the Center for Health Measurement at Duke University.

LISA WEISSFELD: Lisa Weissfeld. I'm a Senior Investigator at Statics Collaborative.

MALLORIE FIERO: Great. Thank you very much. I look forward to our panel discussion later on today. And just as a reminder, this example will be an oncology example. However, we want to think about how the estimand framework can apply to any COA objective. So this is why our entire set of panelists here does not -- is not necessarily in oncology ranges across the therapeutic areas.

Okay. So what is an estimand? So the goal of the estimand framework is to provide

transparency in what is being estimated. Okay? So the definition of an estimand is the target of estimation based on your question of interest. Okay? So basically what am I trying to estimate and explicitly define what you are estimating based on this question.

The draft addendum identifies four attributes that make up the estimand. Okay? So the first one is the population. So the population are the patients that are targeted based on your scientific question. This could be something like maybe you are interested in efficacy population or perhaps I have more of a safety objective. In that case, maybe I'm more interested in a safety type of population.

Next we have our variable or endpoint of interest. And this is the measure that is required to address this scientific question of interest. For a COA objective, this would involve defining the tool, the score, perhaps the type of endpoint if we're interested in being changed from baseline if we're interested in proportion of patients that deteriorated

in say pain, or maybe we're interested in a timed event outcome.

In addition to that, thresholds and estimates that are important in interpreting clinical relevance are also important to specify up front.

Next we have intercurrent events. And this is a relatively new term that was introduced in the estimand framework. And an intercurrent event is an event that can occur after randomization that can impact interpretation of your results. Okay? So an example of an intercurrent event is say if a patient is on a trial and then they discontinue treatment. How does that impact the rest of their COA measurements say for pain? Or perhaps a patient moves on to initiate subsequent therapy. How does that impact how they will score their COA? And so anything after these intercurrent events, it can impact the interpretation.

So it's important to up front in your multidisciplinary discussions list intercurrent events that might be important in your trial and then list how you would want to address them in your analysis.

And that would be in alignment with your question of interest.

So lastly, we have our population-level summary. And this is the basis for comparison. So perhaps we have a randomized trial, we have two treatment arms, and we want to compare the two treatment arms. Well, what summary measure are we using to compare these two treatment arms? This could be something like a difference in mean change from baseline or say a difference in proportion of patients who deteriorated in physical function. And so it's important to also explicitly state what the summary measure is.

So you can see that by specifying all four of these attributes, then it provides transparency to within the multidisciplinary team of what's being estimated. Because these are not new terms. Or they're not new ideas I should say. Right? Statisticians are already making a decision on what analysis population to use. We're already making decisions on how to deal with intercurrent events. It's just that we need to specify them up front and we

want to think about them together as a team more and think about the assumptions that are being made when we have questions.

And one last thing I would like to mention is that there's one final attribute, which is treatment, that was just specified in the final addendum of ICH E9(R1). And this just went live on Wednesday. But for the purposes of today's discussion, we will just go over these four attributes. Okay.

So this figure basically shows how the estimand attributes are placed in context of the research objectives, the analysis, and communication of results. So what I will do is I will go over a case study and then we'll go over each of the estimand attributes, and then we'll stop there. Okay.

So just as a disclaimer, we are not endorsing any particular design, endpoint analysis, but rather we are emphasizing how to think through your research questions. Okay.

So now we go into our clinical case study scenario. So in our scenario, we have a

metastatic or advanced breast cancer trial. And these patients have already progressed out of first-line therapy. So this is a second-line setting. So for patients with breast cancer, they have heterogeneous disease symptoms. And many of them are asymptomatic at baseline, or they don't have symptoms.

So second-line prior studies have shown that -- it says median overall survival of two to two-and-a-half years. So this means that they have a fairly long survival of two to two-and-a-half years median with second-line therapy alone. And that median progression-free survival is 10 to 12 months. Progression-free survival is the time from randomization to either date of death or disease progression. And so OS and PFS are common efficacy endpoints that you would see in oncology trials. So the treatment goal here is that the addition of our targeted therapy will improve PFS by six to eight months.

So we have a randomized control trial. In the treatment arm we have standard of care plus an oral targeted investigational agent. And in our

control arm we have standard of care plus placebo. So for our efficacy endpoint, our primary efficacy endpoint here is progression-free survival. And we expect there to be an improvement of six to eight months in the treatment arm compared to the control arm.

Secondary endpoints will include overall survival. So how long will the patient survive from randomization? However, this can be impacted because patients might initiate subsequent therapy as these patients tend to live fairly long.

So the COA measure that we are looking at here is a physical function score using -- and we're already assuming that we have a well-defined measurement tool, and we're saying that we're going to collect them at every treatment cycle. So perhaps this is every 28 days.

So one thing to note is that in our multidisciplinary discussions, we noted that physical function was an important concept for patients. And that is the reason why we decided to include this in our hypothetical trial. And another thing to note is

that since patients are expected to live fairly long, we expect most of our patients to be on trial at the time of our analysis, which you'll see in a bit.

Okay, so lastly, we have expected safety, symptomatic toxicities. We'll include diarrhea, fatigue, and rash on the investigational arm.

Okay, so now that we have gone over our scenario, the first thing we're going to do is define a research objective. You can see that the rest of the panels here are gray. And we're going to go through each of the estimand attributes and then we'll stop there for the purposes of this presentation.

Okay, so we have defined our broad COA research objective, which is to evaluate efficacy related to physical function. So perhaps we are interested in seeking a labeling claim using physical function. And in this case, we would want to compare treatment arms. For efficacy we want to show that perhaps the invitational arm is superior to the control arm. And in this case we would have to pre-specify a hypothesis and make sure to include them in

our statistical analysis to adjust for multiplicity and testing for multiple endpoints.

So we had many discussions on different types of research questions for this example. And we came up with this one for now, which is, is the average change in physical function from baseline to week 28 better or superior in the investigational arm compared to the control arm?

So as I mentioned, since we're looking at efficacy, we're specifically stating that we expect or we want to see if the investigational arm is superior compared to the control arm. So we're comparing the two treatment arms. We are stating that we're looking at physical function. Because, again, based on patient input, they told us physical function is important to them.

Next, we were thinking about the time of analysis. So if you have a specific timepoint of your analysis, it's important to say justify your reasoning for this timepoint.

So in this case, we chose week 28, which is about six months. And in our discussions,

the clinician told us that week 28 is about the time where the effects of the drug in terms of efficacy and toxicity will have equilibrated at this time. So that is why we chose week 28.

So now that we've defined our research objective, we're going to define our target study population. So based on our research question, we defined our target study population to be defined through inclusion and exclusion criteria to reflect the targeted patient population for medical product approval.

So these are the same patients basically that we're using when we want to show that we want to approve for this medical product. So an example is say we want to include all of our patients who are randomized, right? Since we want to compare treatment arms, we want to include all randomized patients regardless of adherence. So here we're defining our target study population based on our research question.

Next we have our endpoint of interest. So based on our research question, we defined our

endpoint of interest. This is going to be a change from baseline in physical function score. We are, again, assuming already that we have a well-defined measurement tool and that we will use measurements at baseline and week 28.

So now we get into intercurrent events. And again, an intercurrent event is an event that can occur after randomization that can impact interpretation of your results. So based on our research question, you can see on the left-hand side we listed several intercurrent events. So we had a few discussions, multidisciplinary discussions talking about what are intercurrent events that might be important in this particular trial. And so from what I heard in the first session, this is a place where you can include patient input in terms of what might intercurrent events are important and maybe how this could impact their interpretation of the endpoint.

So we listed the intercurrent events. And then on the right side we listed how they will be addressed in our analysis. And they need to be addressed in alignment with your research question.

So first we listed patients who discontinue treatment, disease progression. If a patient goes on to take say physical therapy, after a patient takes physical therapy, how does this impact their physical function score, right? And then we have subsequent therapy. So how do these events impact how a patient might interpret their physically function score. And it might be positively or negatively impacting the score. But it's important to have these discussions early on.

And so what we state here is that we will continue to collect and include the observations in the analysis regardless of whether these intercurrent events occur.

For example, if a patient say moves on to have physical therapy or disease progression, we will still continue to collect them. And that's because we want to make sure we have the least amount of missing data. And eventually we would want to compare the two treatment arms to avoid bias in the case of any missing data.

So next we had the intercurrent event

of death. In the unfortunate case that a patient dies, we cannot collect their physical function after that intercurrent event occurs. Right? So how is that dealt with in our analysis? And oftentimes I don't necessarily see how this is dealt with in the analysis in a statistical analysis plan for COAs, which can be surprising. But we had many, many discussions on different ways to address death as an intercurrent event in our analysis. And so what I will say here is that it needs to be addressed in your analysis plan, and it may be included as part of an endpoint. This is one option to address death as an intercurrent event.

But as I mentioned, our clinicians stated that we don't expect a high proportion of patients to die at the week 28 timepoint because they are -- the patients are expected to live fairly long. So we don't expect many patients to die in this case.

Okay, so lastly, we have our population-level summary. So based on our research question, we have our population-level summary. And if you remember, our endpoint was mean change from

baseline in physical function score. So our population-level summary is the difference between treatment arms in mean change from baseline in physical function score using baseline and week 28 measurements.

Okay, so this is our summary slide of say all of the decisions that we made for this particular COA objective. Okay? So we are explicitly defining the target population, the endpoint of interest. We thought about intercurrent events as a team. We thought about how to address them. And I also wanted to note that you don't have to address all of your intercurrent events in the same way. In this case, you can have say two different ways that you're addressing your intercurrent events, but that should be pre-specified up front. And then we have our population-level summary.

And so again by doing this, we are providing transparency between our multidisciplinary teams as well as the regulators and industry in terms of what exactly is being estimated. And this will help us to improve our interpretation of COA results.

And so this is a final slide just showing you, again, the figure of how the estimand attributes are placed in the context of the research objective, statistical analysis plan, and communication of results. Excellent.

So now we will move on to our panel discussion. So I will have two panel discussion questions. And we will make time for audience Q&A. So start thinking about any questions that you might have. Since lunch starts at 12:00, we'll have audience Q&A at about 11:50. Okay?

So the first question that we would love to hear from our panelists is what do you foresee as real-life challenges when using the estimand framework for a COA research objective. And in addition if you could please discuss any considerations in addressing intercurrent events. And I'd like to remind our panelists that this question isn't specifically for the case study that we just talked about, but rather it's for any COA objective. And if you could limit your speaking time from one to two minutes, that would be fantastic. And we will

start off with Lisa and then we will move down the line.

LISA WEISSFELD: First of all, I'd like to point out what I think the advantage of this approach is. And that is that it provides a framework to have a discussion around the study design and it sets up essentially a common terminology for the group, the team to discuss the issues. And so there are real advantages to having a framework. There also can be disadvantages in that they can be -- you can treat them in a very prescriptive manner and limit your options. But the real advantage is that it's a framework. And that framework invites people to the table to have a discussion using common language. And I think that that's the single most important contribution of this framework. And also, to make it a little simpler, it's like who are we studying and what are we interested in. When we give patients a treatment, what is the real effect? What's the effect that we are going to measure? And then how as statisticians, since I'm a statistician, my focus is on summarizing the information. And so how at the end

of the study will we summarize that information and how will we do that in a manner that fits on a label nicely, but also has meaning to both the clinicians and the patients.

With respect to intercurrent events, I think that that is a really difficult problem. And it's something that we all are faced with when designing study. And in the particular example there that was presented, the intercurrent events were all grouped together. And one of the challenges I think going forward is that those vents are not equivalent. And so when we have events like disease progression versus some of the other events that were on that slide, is there a way to treat those differently within the framework? How do we go about doing that?

MALLORIE FIERO: Thank you. Kevin?

KEVIN WEINFURT: I agree with Lisa too that the framework is so helpful. And as someone who is at the table lots of times with clinicians, biostatisticians and others, and patients who are giving their perspectives, it's a really nice say of organizing those discussions.

I had two specific areas where it seemed like there might be some opportunity for clarification though. And they both touch on intercurrent events in a way, and they both deal with multi-component endpoints.

So one of the great things about considering that class of intercurrent events that are things people do to try to make themselves feel better, like taking some extra drugs or use some assistive device. When that's happening and I want to be measuring degree of symptoms of functioning, I've got this interesting situation. And the guidance rightfully points out, well, one thing you could do is to decide to bring those intercurrent events into the endpoint as part of the endpoint definition. And that's terrific.

As the discussion document stands right now, there is a little bit of a tension created. Because if I've got some continuous or ordinal symptom or functioning scores and I've got some type of indicator of that thing the person did to make themselves feel better, I want to bring them together

to define a multicomponent endpoint. One thing I might come away from the discussion document with is FDA doesn't like responder criteria using thresholds in the definition that will be used for the hypothesis testing. And so then I'm trying to figure out how would I put that multicomponent endpoint together if I can't do that? What are some of my options? So that's one thing about the multicomponent endpoints.

And the other thing the document rightfully points out too that one of the considerations for multicomponent endpoints is the weighting of the individual components. And it uses the term reasonably similar clinical importance at one point as one of the statements. And as we look across guidances, look at the draft guidance for multiple endpoints, there there's reference made to the weighting, but there the language is clinical importance (indiscernible) is substantially different.

And so I, as a researcher and a user of these, might be wondering, well, what would reasonable approaches be for arguing for a particular weighting scheme? I know that it's a consideration, but what

type of things would I bring to bear to make an argument for one or another weighting scheme. But other than that, I thought it was extremely helpful.

MALLORIE FIERO: Thank you.

DANIEL SERRANO: I'd like to agree with everybody else and also echo back to Fraser's point in the first panel about these being very useful points to begin conversations with.

I guess when I think about some of the potential challenges to implementing this, especially thinking about intercurrent events and a related concept. I think one of the things that's kind of implicit and in some cases kind of very explicit in this document is the idea that post-randomization -- right? Randomization does a very good job of kind of wiping out other-worldly effects. And any of the things that we need to consider are the intercurrent or intervening effects that would occur post-randomization.

But I think even the practice effect example kind of speaks to the fact that things that can occur say in a run-in period prior to

randomization can propagate into and influence the post-randomization phase.

So while I think it's very important to think about all the intercurrent events or intercurrent effects that could crop up and how to deal with them, I think there are also potentially in some cases very important and meaningful intercurrent effects, right? We saw this in several trials we were working on last year.

One of the ones that I think crystalizes the most was a really interesting migraine trial. And there were two double-blind periods. And subjects who were randomized in the DB1 were then re-randomized in the DB2. What we found in doing some sensitivity analyses was that the single best predictor of response in DB1 was the subjects' report of their typical response to their typical migraine treatment outside of the trial, whether you're in placebo or treatment. When we went into DB2, the single best predictor of response in DB2 was response in DB1.

So I think the intercurrent effects

framework is very useful for thinking about potential problems and how to kind of address those that could disrupt our ability to (indiscernible) the effective interest.

But I think to implement this effectively, I think probably viewing this as a kind of starting point for discussion, perhaps we can also use that framework to then step back out and say, well, are there other individual variables that could come into the trial and are not really mitigated effectively by randomization.

JOHN SCOTT: I think those were all good and useful points. Also I'm obligated to say that Lisa's points are good and useful, because she was my professor. So I'm sure there's some way she can go back in time and fail me for inference.

I think the main challenge for implementing the estimand framework is really the same for COA and for hard endpoints or non-COA endpoints. And it's two things. One is that there's a lot of careful thought that's required for going through each of those four or five things. People are kind of used

to sort of default analysis sets, giving a little bit of thought to a missing data technique and then you're done. In the estimand framework, you're really called upon to think very hard about things. And that is a challenge.

The other challenge is very real-world, which is the change management to get people to feel like they need to do something new, they need to use the estimand framework. So those are the main challenges.

There is one area where I think COAs present a greater than average challenge, which is specifically the population-level summary element of the estimand definition. So if you're working with something like mortality, it's quite clear that you can use a difference in proportion of surviving patients as an endpoint or as a population-level summary. It's unambiguous what that means.

If you're using a symptom score and you want to use for example difference in mean symptom score between groups or difference in mean change in symptom score between groups, then you're asking

questions like what does a mean difference of 1.7 mean. Excuse me for repeating mean.

And so that can be a very difficult question to address. And for many COAs, 1.7 doesn't necessarily mean the same thing at every point in the scale. So that's worth considering. Also, people are sometimes tempted to create binary endpoints out of underlying continuous COAs. I think that tends to make the interpretation even worse, even harder. And it definitely makes the statistical power worse. So it requires a lot of extra thought for these endpoints.

GREGORY LEVIN: So I have two comments, and I think that they are for kind of any objective and not just the COA research objective, but I'll try to give some examples to illustrate them using COAs.

The first one is that I think to do this well, the interdisciplinary conversations, the relevant input, including from patients, needs to happen early and often. And the first reason is the reason that John said, which is that this is hard. And the second reason is that -- I like this question

because it says a COA research objective, it doesn't say the COA analysis or the COA endpoint. This is about the objective which needs to be determined before you come up with a design.

So I think one of the things that we've seen a lot is that the protocol is already established, design is already established. Maybe even the sponsor is at the stage of submitting a statistical analysis plan, and that's the first time that the estimand framework is being discussed and brought to us. And that's too late.

I think the ultimate goal of this is to think about things like the control arm, like how background standard of care and ancillary medication is handled in the trial, whether patients are followed or how they are followed after certain intercurrent events. Those things need to be determined early.

So just to give an example with an intercurrent event, if there's a thought to be an interest in a treatment policy strategy that there is some interest in knowing, say, a patient's function even after they discontinue treatment, the design

needs to incorporate strategies to follow patients and assess their physical function even after they discontinue treatment. That's something that needs to be built into the design and the protocol.

The second I guess challenge is I think trying to answer too many questions at once, trying to think that we have to answer every single question and address every single objective in a single analysis -- I mean, I'm a statistician; I think the primary analysis is important. But I think the addendum talks about something called supplemental analyses or supplemental estimands. And I think that's something that we should pay a little bit more attention to.

An example of that is the one that John just mentioned, which is you could have a primary analysis that looks at effect on a mean, but there could be important supplemental estimands or supplemental analyses that address other summary measures of the distribution such as looking at the cumulative distribution function or looking at the proportions of patients that meet certain thresholds. Those are answering slightly different questions, but

they can be important to understanding the benefits and doing a benefit-risk assessment.

So those are -- I guess one other comment on that is that -- another example of that is when you have a composite strategy where you potentially consider a patient who discontinues treatment or uses a rescue medication as say having a poor response. You know, that kind of an analysis, as with any composite strategy, you can have a drug effect driven by an effect on adherence alone.

And so something like a supplemental estimand where you look at functional improvements regardless of whether patients discontinue treatment can be important.

So I think the challenge is paying a little bit more attention to the importance of supplemental estimands and supplemental analyses to better understand benefits.

JESSICA LEE: I guess I'm the only clinician on this table. So I agree with everyone, what has been said so far. I guess from a clinical standpoint, one of the major challenges that we face

is that we don't always have the best understanding of the natural history of the disease for many diseases, even the diseases that we think we know. When we speak with patients, we realize that we may not be targeting the symptoms that really matter to these patients. And this is particularly challenging in rarer diseases where natural history is not very well known. But even for common diseases that our division encounters -- and I predominantly oversee GI applications -- we are struggling with trying to figure out an endpoint that's going to be assessed in a heterogenous population with various different symptoms and a disease course that's not always predictable.

And so we're always faced with what are the COAs that need to be collected to help inform that endpoint so that the data will be as generalizable as possible to the intended patient population and how frequently and how long these assessments need to be made so that we can get meaningful and interpretable data at the end of the day. And we also have to be mindful of the patient burden of taking these daily

diaries.

In our division, we actually frequently have patient listening sessions to try to understand what is the patient burden and what are the signs and symptoms that are most important to them and what would they consider to be a meaningful outcome.

And, you know, I've taken care of patients, but it's not always the same what the physicians think are important versus what the patients feel is important. And I'm not saying one is -- you know, both of those are important. But really until you have that dialogue, it's really hard to know for sure.

And I'll just give a quick example about the intercurrent events. In GI because we're often faced with these chronic diseases with relapsing-remitting type of nature and with heterogenous presentations, that we usually request long-term clinical data to help characterize the efficacy as well as safety to inform chronic administration.

But the requirement to collect daily

data from patients for a large proportion of the trial duration can be very burdensome for these patients and can result in a lot of missing data. And there's definitely a high likelihood of having intercurrent events such as recue medication use that needs to be taken into consideration when we're thinking about the estimand framework.

So I am very fortunate that we have great statisticians and COA staff that we work with. And I have to say in our GI diseases, they're in almost all of our meetings and we have these very frequently. So I'm actually very excited to see this document where we're trying to explain all of the things that are very -- that we struggle with on a day-to-day basis.

MALLORIE FIERO: Thank you very much. There were a lot of good points that were brought up by the panelists. And a few of the things that I heard was that the framework provides a common terminology and that there's a lot of thought that is required for us, and the estimand framework helps us to do that more so. And one of the things that you

mentioned, Jessica, was that you have frequent discussion about it. And I do think that's very important, to have frequent discussions and early discussions about the research objective, the design, analysis, and interpretation. So in the interest of time, we will move on to our next question.

So for question two, how does a treatment's mechanism of action, disease's natural history, et cetera, impact steady duration and timing and frequency of assessments for COA endpoints. And I know that Jessica kind of touched base a little bit on that. And since she is our resident clinician, I would like to start with her first, and then we will move down the line. Jessica?

JESSICA LEE: Okay. So I actually think that those are probably two of the most important factors that go into help determining the study duration and timing and frequency of assessments for COA endpoints. So it's critical to understand the natural history, to identify the patient population that's most likely to benefit from a given treatment, especially based on its mechanism of action and

endpoints that would be most meaningful for these patients. And how frequently and how long COA assessment should be made could be informed by natural history as well. And this is largely informed by an understanding of the disease course, whether it's acute versus chronic, is it slowly progressive or a rapidly-progressing disease, is it a relapsing-remitting disease, is it episodic, or is it mostly stable? All of these things are important to understand for us to figure out how long should this study be and how frequently do we need to assess these COA endpoints.

And the drug's mechanism of action is especially important because that's going to help determine what is the aspect of the disease or key signs and symptoms is the drug able to target. And this is particularly important for disease that present with heterogenous presentations where you may not be able to target all of those signs an symptoms with one drug. And time of efficacy assessment and trial duration is also affected by the type of drug that you have, because they all have different

expected onset of action. And when the drug is likely to take effect on the effective outcome of the interest could differ depending on the mechanism of action of the drug.

MALLORIE FIERO: Thank you. Greg?

GREGORY LEVIN: Yes. I guess the quick answer is that it impacts it a lot. And to illustrate that, I guess I can just talk through a specific example. And so that would be the difference between say an acute pain setting and a chronic pain setting. So suppose we have a COA that is a patient-reported outcome assessment of a patient's pain perhaps with a numeric rating scale, either asking for a patient's average or worse pain in the last 24 hours or maybe asking about a patient's pain now.

We take that COA or some variation of that COA and we put it into a chronic pain setting. Maybe a drug intended to treat patients with say knee osteoarthritis or chronic lower back pain. And the goal there would be to have a treatment that has chronic benefit on a patient's pain. So long term, durable improvement in their pain over time.

And so to evaluate that research question, you would often want some duration of a trial to have a reasonable kind of surrogate for longer-term benefit. I think the approach is often to use something like three months. And the endpoint would often be an assessment at the end of the treatment period. Maybe an average over a week, with the idea that -- for example, if you have a patient who has a very short-term benefit in their pain but then can't tolerate or adhere to the drug or the effect of the drug goes away, that wouldn't be a very good outcome for their chronic indication. So that's the goal of perhaps using something at the end of the treatment period as an endpoint.

On the flip side, if you had an acute pain treatment maybe post-surgery, surgery for bunionectomy or hernia or something like that, the goal is short-term improvement in the patient's post-surgical pain, you might only need a trial of a week or certainly an endpoint that captures a patient's pain over a few days, which is really the intention of the treatment and the goal of the treatment. And that

might be something like an area under the curve. They often use something called SPID, which is essentially the area under the pain curve over a few days. So that's just an example of how both the nature of the expected effect of the drug and the disease's natural history can greatly impact things like the duration and the timing and the choice of the endpoint even with a similar underlying instrument.

JOHN SCOTT: My answer is very similar to Greg's, so I'll keep it brief. I agree that these considerations are extremely important for trial design considerations, including duration and timing of assessments. I have a similar acute example. I think mine is even shorter-term.

At CBER, we have approved multiple C2 esterase inhibitor products for acute attacks of hereditary angioedema. And the way these studies work is that patients come into the clinic or to the hospital in significant distress. They are given treatment. And then maybe every 15 minutes or every half hour they are asked to self-assess how they're feeling. And the endpoint is how quickly they get

relief from those attack symptoms.

So the whole trial for one patient lasts about 24 hours. So what that's based on is understanding that the mechanism of action of the product, it's a therapy that should have a very quick effect. And the natural history is that these attacks are somewhat self-limiting anyway. So if you haven't responded in 24 hours, you're sort of out of the window for an effect.

On the other end of the spectrum with a lot of our gene therapy products, you're looking at a situation where it may take a long time for a new protein to be produced and then for that to have a downstream clinical effect.

For the gene therapy that was approved last year for a rare form of childhood blindness, really the only endpoint assessment, the primary endpoint assessment was one year after treatment because it was understood that it would take a while first of all for the treatment to have an effect. And second, you wanted to make sure that it had some kind of lasting effect.

DANIEL SERRANO: So I think the panelists have to my mind, at least for my considerations, addressed the general question here.

I guess I would take this then and I would extend it just one step to the estimand. I think there are definitely implications here for these sorts of things and the nature of the estimand that is selected for a given efficacy assessment.

A brief example. You know, what's something we see in the context of say prophylaxis trials for chronic diseases with episodic attacks or relapsing-remitting is you can have inclusion-exclusion criteria where you do a run-in to kind of make determinations of the subject. It's like sufficiently symptomatic at baseline prior to randomization. And in a lot of these diseases like that where you kind of have these kind of cyclical episodes, you kind of effectively roll them at zenith. And then what comes next, even in the comparator or placebo arm, is nadir. And so when you think about deployment of a change from baseline estimand in that context, you know, you can very easily run into

trouble. And so thinking about the natural history of the disease and then how we're going to kind of quantify the benefit of effect in that context, you can think of alternatives in this kind of estimand framework. For example, you know, minimizing volatility. Right? Making someone's day-to-day life more predictable and that sort of stuff. Right? And that kind of may be a greater or more sensitive framework than say the change from baseline.

So I guess all I'm saying is that I think natural history and the nature of how this disease is going to function just independent of intervention is probably likely going to affect things beyond simply the administration of the schedule of assessments, but what endpoint or what estimand you're going to end up using or will be most effective for you.

KEVIN WEINFURT: I don't have anything to add to my colleagues' comments.

LISA WEISSFELD: Yeah. I think people did a great job, actually, going through the possibilities. The only thing that I would add or

reiterate is knowing the natural history is really important. And particularly when it's a slowly-progressing disease. And in the rare disease space when you have a slowly-progressing disease and you're trying to design a study, that is incredibly challenging to say the least. And also the acute versus chronic is also another -- these are nice categories. Statisticians like to categorize things. So these are also two fairly distinct categories of problems with the acute oftentimes being easier because you have the subject captive during that acute phase and you have less missing data to deal with than you do when it's chronic. But anyways, so that's it.

MALLORIE FIERO: Thank you. So just as a follow-up question -- and I will ask my FDA statistical colleagues first -- if we moved back to the example that we have. And we had our target population as defined here based on inclusion/exclusion criteria to reflect the targeted patient population for approval. So my question for you is is this -- will this always be our target population? What are some considerations in thinking

about your target study population?

JOHN SCOTT: Thanks, Mallorie. So I think this came up in our discussions earlier, and it's an important question. There's sort of different levels of target population as you sort of expand through the drug development timeline. When you're talking about a phase 3 clinical trial, there's typically a very large number of inclusion and exclusion criteria which are there for several reasons. They're there to protect study subjects, they're there to enrich the population, to get patients who are more likely to respond to the treatment. And they may be there for logistical reasons as well.

And so assuming a trial is successful, some subset of those inclusion/exclusion criteria then typically go on to define the population. So the indicated population is not exactly the inclusion/exclusion criteria. That's always much more specific.

And then the question is once it's approved, how is it going to be used in the real

world? How is it going to be used maybe a little broader than the indication but still more or less as intended. And then further to that, how is it going to be used off-label? So it's complicated questions. When we're talking about trial design, we are typically targeting what the indication and the product labeling will look like. But all of these things come into play.

MALLORIE FIERO: Thank you. Greg?

GREGORY LEVIN: I agree with all that. And I'll just add that I think in some cases you may have a primary estimand that's looking at the indicated population or something more closely reflected to what's in the inclusion/exclusion criteria. But you may, for example, in some benefit-risk assessment discussions also consider the expected use of the product in the real world which may go beyond, as John mentioned, what's the indicated use.

MALLORIE FIERO: Great. Thank you very much.

Okay, great. So this concludes our pre-set discussion questions. I will now open it up

to the floor for audience Q&A. Please direct yourselves to the mics in the middle of the aisle. And we ask that you please be specific with your question and limit your speaking time to 60 seconds. Okay, great. Thank you. Yes.

LISA KAMMERMAN: Hello, this is Lisa Kammerman. I just want to comment that this framework is very helpful. Having been a reviewer at FDA and having worked in industry, I think we can make this framework a little bit more specific. And to build off of some of Greg's comments, the scientific question needs to be drilled down even more.

So, for example, among patients who are still on study drug at week 28, or is it all patients at week 28 regardless of what happened during the study? So that's an important distinction to be made.

And I have difficulty with the word handling. And it comes down to a distinction between the protocol and the statistical analysis plan. I think a better word when we're talking about a protocol design is actually collect. So what to collect really depends on the research question. So

if we're interested in the treatment policy, then obviously we have to collect all the COAs up to week 28. But if we're only interested among those who stayed on treatment, then we don't need to collect events afterwards.

But when we get to the analysis plan, the word handle really gods into the analysis. How are we going to handle or treat these events in the analyses. And there is a common kneejerk reaction I think, well, let's just MMRM the data. But really it should be the other way around. Are we going to assume that deaths are missing at random or are they going to assume a value of zero, for example, or something else? And then you use that to define your analysis rather than saying let's MMRM the analysis.

MALLORIE FIERO: Thank you. Yes, John, please.

JOHN SCOTT: Lisa, those were great comments. I wanted to sort of second the idea that talking about handling intercurrent events is not necessarily the right psychic frame of mind to be in. I think the real question is how do the intercurrent

events, how are they reflected in your scientific question. Right? The intercurrent events change what question you're trying to answer. And so they're not a nuisance; they are core to what you're trying to do.

MALLORIE FIERO: Great, thank you. I think in the back, the very back. Yes.

GRACE WHITING: Thank you. I'm going to start my timer so that I don't talk over.

MALLORIE FIERO: Thank you very much.

GRACE WHITING: So I am Grace Whiting. I am the president and CEO of the National Alliance for Caregiving. And I thank you for hosting this meeting. Thank you, FDA, for what we observe to be an increasing awareness of the role of the unpaid family caregiver.

I just want to raise the issue though that when we talk about clinical outcome assessments, it's not very well-defined when a caregiver may step in and provide information either as a proxy for someone with cognitive impairment or in the pediatric population. And in other cases where the caregiver may by default be providing information. You see this

in clinical settings all the time. For example, a family member may be the one that actually logs on to an EHR and puts in information or collects information or coordinates care.

So I don't know what the question should be, but I'm just raising the issue that there are unpaid caregivers across the lifespan that are a part of this process and that the relationship they have with the person receiving care can sometimes impact the way they report data out.

MALLORIE FIERO: That's an interesting point. Do any of the panelists have any comment on that? Thank you very much.

Okay, so moving on to the person in the middle. Thank you.

ANDREW TRIGG: Thanks. So I'm Andrew Trigg. I'm a statistician at Adelphi Values, a consultancy. And so one of the points was around kind of considerations into current events. And I think for me one of the big things is whether the occurrence of that intercurrent event has a kind of causal link with treatment.

For example, we talk about kind of rescue medication and kind of -- so, for example, for pain. If you're on a poor treatment or control, then your pain may be getting worse. So your use of rescue medication could increase. Whereas I know in the discussion document we have something around -- I think it's like if a patient broke their leg and their physical functioning scores would be worse. Which, you know, that kind of wouldn't be related to the treatment causally.

And so I guess kind of a thing to think about is if the intercurrent event is not causally related to the treatment, do we need to think hard about controlling for it, or do we just consider it as kind of part of the random error inherent in our measurement?

MALLORIE FIERO: Okay. So what I heard is if you have an intercurrent event that's not directly perhaps related to the treatment, what are things to think about when thinking about how this event can impact your interpretation, how can you deal with this. Any thoughts from the panelists? It's a

touch question.

LISA WEISSFELD: We had something recently in a study where it was a physical function that was being measured. And there was one subject who broke their ankle. I mean, the same example. And I think what we ended up doing was until the ankle healed, we did not include those values because they were missing data. But then when we were doing the sensitivity analyses, we did impute those values.

MALLORIE FIERO: Greg?

GREGORY LEVIN: I think that's a really good question. I think the only thing I would say is that I would agree that the conversation is much more difficult about what you are trying to estimate and what the question is when you're talking about an intercurrent event that is plausibly related to the treatment assignment.

MALLORIE FIERO: Great. Thank you very much. In the front, please.

ROSS WEAVER: Thank you. Ross Weaver with Clinical SCORE. I have a simple question. When using the estimand framework, is there a role for

listening to the patient voices beyond just understanding the natural history of the disease, and if so, when during the process and where? Phase 2, phase 3? I'm just trying to understand where it would fit in.

MALLORIE FIERO: Yeah, that's a really great question. So how does patient voice fit into the estimand framework? Thoughts from the panelists? Jessica?

JESSICA LEE: I guess I can start. As soon as we start even getting applications, we actually start getting -- if we don't have enough information internally and we would reach out to patients to get their input. So I don't think it's a specific timeline, but I think as soon as possible as we start planning for it. Because I think we've heard from everyone that it's a lot of work and we need a multidisciplinary team And it's important to get patient input as early as possible and as frequently as possible.

JOHN SCOTT: Yeah, it is a very good question. I think patient input can help in every

stage of defining the estimand. The two stages that I think are particularly critical are defining the endpoint of interest; what is important to the patients. And also again back to what I was saying earlier about the population-level summary, what information can the trial provide that will be meaningful and will help people make medical decisions for themselves and their loved ones.

MALLORIE FIERO: Great, thank you. A question in the back?

CINDY GIRMAN: Yes. Cindy Girman, CERobs Consulting. Thank you for this session. I really enjoyed it. First, I was glad to hear that it was the effect on the intercurrent events on the analysis and interpretation. And that speaks to sensitivity analysis, and I think it's really important to emphasize that you should also be prespecifying the sensitivity analyses. Otherwise, there's just a lot of analyses that are done.

And second, I think you still have to prespecify a main analysis. And I wonder if it would be helpful to give some examples where you may have

non-trivial proportions of patients dropping out for AEs or for ineffective therapy. In other words, not missing at random, whether you should do a rank analysis where you're considering those to be the worst values, or if there are other approaches that would be suggested.

MALLORIE FIERO: That is a very tough question. And I know that at least for the statisticians at FDA, we think about these issues all the time. I don't have a specific answer for you, unfortunately, but just that it's important to think -- I think this is a real problem. Particularly for me, since I work in oncology, a lot of times there's issues with patients who, say overall survival, if patients die, it differs by the treatment arms, and how do you deal with this in our analysis? And this is not something that I know the answer to, but that is definitely something that we need to think about. And sensitivity analyses are definitely important.

Were there any other further comments on that? Okay, great.

I will take one last comment. In the

front.

ALYCIA SHILTON-LLOYD: Hi. I am Alycia Shilton-Lloyd from Gilead Sciences. My question is really around -- there was some discussion in the guidance about the different modes that you use to collect the information on the tool. And I think I'm a little bit confused by the discussion then on making sure that the flexibility exists around intercurrent events to capture them as needed when the modes themselves, there are very limited times that you can think about using multimodes. So it might be helpful to have some examples of what that means to the guidance. What are some of the examples of when you can shift modes, both from two perspectives. The first is if you have sort of a decline in function of in patient activity that limits their ability to -- you know, more than just the broken ankle. But if you are taking input in vocal assay. So verbal assay using natural language processing or something similar and there is an unexpected decline in a significant number of your population in breathing ability. So that they purposefully choose shorter words in giving

their input. How would you adjust modes if it's unplanned for? You might not have seen it in a smaller population. And then second, because we're insisting on a specific mode, are we limiting in some ways who can participate in the trial if they communicate in a different preferential mode? So having some examples around what it might look like to have multi-modes would be helpful.

MALLORIE FIERO: Thank you very much. You raised a very important point, and I think definitely please submit that comment into our public docket so that we can address this comment.

Do I have any further thoughts from the panelists before I close this out? Okay, great. So I wanted to thank our panelists and our audience participants.

Just as a recap, I have a bajillion notes that I wrote here. So a few of the things that I heard is that the estimand framework is important because it provides common terminology for our multidisciplinary teams to talk about issues. We need to think about intercurrent events and

multicomponents, that there are also events and other variables that can happen not necessarily after randomization that may impact interpretation. A lot of thought is required for us in terms of thinking about the COA objective, and we need to make sure we stand up to this challenge. And there are challenges in a COA score that we can't just say make it a binary endpoint. It will be difficult to interpret. So we need to think about interpretation of our COA endpoint. That input of patients needs to happen early and I think at every stage of our study. Supplemental and sensitivity analyses are very important, and it's a challenge to understand your natural history, so it's important to discuss that with your multidisciplinary team.

I also heard that the two most important factors are treatment's mechanism of action and disease's natural history in determining your trial curation and timing and frequency of a COA endpoint assessment. And we heard a lot of really great examples of how these can impact trial duration and timing of assessments.

And so I'd like to thank our panelists for their thoughtful comments. Just as a reminder, I'd like to encourage everyone to comment on the discussion document. And the public docket will close on February 4th.

I would also like to thank Madeline Pe from EORTC in (indiscernible) and Chana Weinstock, who is our oncologist at FDA, and many others who were involved in developing the example that you just saw today, which is in Appendix 1 of the discussion document.

We are now going to enter into our lunch break, yay. Session 3 will begin at 1 PM. If you could please give a round of applause to our panelists in this session. Thank you.

(Break)

LILI GARRARD: So welcome back. I hope everyone enjoyed lunch and also has had a chance to catch up with friends and colleagues.

So my name is Lili Garrard. I am in the Office of Biostatistics, Office of Translational Sciences, in the Center for Drug Evaluation and

Research here at FDA. And I will serve as your moderator for Panel Session III: Considerations When There Is Heterogeneity in Disease Symptoms and Functional Status Between Patients and Within the Same Patient Over Time. And we have a great panel ahead of us today, and I really look forward to the exciting discussions that we will have.

As we have heard from earlier discussions, this is really a challenging area for a lot of us. And we all have a lot of thoughts and questions, but we don't have optimal solutions. But this is why we're here today, gathered around to talk about the consideration and the challenges.

So before we start with the panel discussion, I would like to first have my panelists just introduce themselves. So, Lisa?

LISA KAMMERMAN: Hi. I'm Lisa Kammerman. And this is the first time in my life that I've been at the top of the alphabetical list.

[LAUGHTER]

Oh, I'm sorry. I'm a biostatistician. I was a reviewer at FDA for 24 years, and have worked

in the industry, and now I'm an independent consultant.

ELEKTRA PAPADOPOULOS: Hi. I'm Elektra Papadopoulos, and I lead the Clinical Outcome Assessment Division and the Office of New Drugs here in CBER.

TEJASHRI PUROHIT-SHETH: Good afternoon. My name is Tejashri Purhohit-Sheth, and I'm in the Office of Tissues and Advanced Therapies in in CBER, where I'm the Director of the Division of Clinical Evaluation and Pharmacology Toxicology.

DAVID REASNER: Hello. I'm David Reasner. I work at Imbria Pharma. And I'm a psychologist by training, but have spent several decades developing medical products, working primarily in biostatistics and COA development.

STEVE ROBERDS: Good afternoon. I'm Steve Roberds. I'm the Chief Scientific Officer at the Tuberous Sclerosis Alliance, which is a non-profit advocacy organization based in downtown Silver Spring, not too far from here.

PATROULA SMPOKOU: Good afternoon. My

name is Patroula Smpokou. I'm in the Division of Gastroenterology and Inborn Errors Products. And I'm in the team that regulates products for inborn errors of metabolism. And that's in CDER at FDA.

R.J. WIRTH: Hello. I am R.J. Wirth. I am Managing Partner for Vector Psychometric Group. I'm a quantitative methodologist and psychometrician by training, and I oversee the day-to-day operations as well at VPG.

LILI GARRARD: Thank you all for being with us today. I think we can all agree that having a well-planned COA strategy is critical to support the selection and interpretation of COA-based endpoints in medical product development programs.

However, one of the major challenges to COA measurement and endpoint construction in clinical trials is the heterogeneity in diseases that all stakeholders often have to deal with.

So do help us kick off the discussion today, I've listed some example heterogeneity in diseases on the slide. And we all know this is a short list and these examples are not mutually

exclusive.

So there are genotypic and phenotypic heterogeneity in many diseases. For phenotypic, reliability may range from monosymptomatic to multisystemic diseases. Reliability can be seen in terms of disease manifestations, the rate of disease progression over time, and baseline severity of symptoms and functional status. Some diseases may also have a waxing and waning nature and may affect a wide age range.

So in a nutshell, we often deal with heterogeneity in diseases between patients and within the same patient over time. Therefore, it can be a real challenge to assess a single concept of interest across all patients.

So, with that said, I have a couple of questions for our discussion panelists today, and we'll start with the first one. What factors should be considered when developing a COA-based endpoint for diseases with heterogeneous patient populations and variable manifestations? And when you address this question, please also include any potential analysis

and interpretation issues that you would like to discuss.

So, to begin the discussion, I would like to first pose the question to Steve, coming from the patient perspective. Steve?

STEVE ROBERDS: Thanks. I thought I would take just half a minute or so to describe tuberous sclerosis complex, because that's the lens through which I guess I'm going to describe perspectives on heterogeneity. But there's a lot of heterogeneity among lots of different diseases. So that's why I wanted to describe this so it kind of puts it in perspective.

So tuberous sclerosis complex is a rare genetic disorder. People with TSC develop epilepsy, autism, tumors in various organs throughout their body at different times during their life. The cause is mutations in the TSC1 or the TSC2 gene, and the gene products form a complex that regulate the activity of mTOR. So much, so good, it sounds fairly straightforward.

But it turns out it's an autosomal

dominant disorder. So that means people who are born with TSC generally have one bad copy of either TSC1 or 2 gene. But at different times during their life, they develop second hits, or loss of heterozygosity in some cells in their body, so determine the occurrence of when and where in the body and during their lifespan these second hits occur. Has a major impact on the phenotype because it's really those cells that lose completely the TSC1 or the TSC2 function that causes the problems for the disease.

So I wanted to kind of -- I think to address your question, an important factor to consider -- and I'll probably leave others to discuss the potential issues -- but is that with all of these different phenotypes in the disease that are caused by a consistent metabolic pathway, genetic cause and metabolic pathway, there's an opportunity to measure different endpoints in different people, but to actually expect a given drug to work on all of them.

So, as I mentioned, TSC1 and 2 regulate the activity of mTORs. So mTor inhibitors are approved by the FDA for the treatment of epilepsy,

treatment of a certain type of brain tumor which is independent of the epilepsy, and treatments of a certain type of kidney tumor in TSC.

These were all done in independent studies, but theoretically, they could be done in one study. It's just that not all of the people with TSC will have epilepsy, not all of them will have the kidney issues, and not all will have the brain tumor. Some will have two of those things and some will have three of those things.

So that's, I guess, a fairly simplistic examples, I think, with regard to TSC that needs to be taken into consideration when bringing forth a treatment If the mechanism of the treatment is likely to affect the main disease process, there are lots of things that it potentially could affect in those people, but it'll be different from person to person.

I know you've got a second question, so I'm going to save a little bit more the discussion for that part too.

LILI GARRARD: Thank you. Elektra?

ELEKTRA PAPADOPOULOS: And thank you

very much. That was a really good introduction and important patient perspective.

You know, in terms of this topic of heterogeneity, it is very broad and the topic itself is heterogeneous. But you know, I think all of this also goes back to the patient. And as a common thread throughout all of the guidance series that we have, from the very first guidance in the series, which was focused on obtaining representative patient input, heterogeneity comes up very heavily in that guidance. And so it is a key starting point, understanding that heterogeneity and getting input from diverse patient populations.

And here at FDA, we have various mechanisms of doing that, from large public meetings to smaller patient listening sessions, even within drug development programs, sometimes sponsors will bring patient advocates to industry meetings. And so there are different ways of getting that critical patient input.

But for today, I think it's a very good idea to drill down in this broad topic two different

types of heterogeneity. And so we heard there is this phenotypic heterogeneity. Even when there's a common biologic source of a disease, manifestations can be very different. So different patients have different symptoms.

Take, for example, migraine, where migraine headache, all the patients by definition have pain, have headache. But then patients can vary in other symptoms, such as sensitivity to noise or light, or nausea, for example.

And so this is an example of the phenotypic heterogeneity in a disease. And there is FDA guidance on this as well of how to approach this challenge. Now, of course, migraine is a common disease, but we also have challenges in a lot of rare diseases as well, which we'll get into in this session.

And then in addition to that, there are numerous other intrinsic patient factors, such as the age of the patient, other demographic factors, comorbid conditions. And all of these things can pose measurement challenges.

And so it can be very difficult, for example, in pediatrics and in rare diseases which affect a lot of children and patients across their lifespans to identify a common measure that can't be used across the age groups.

And yet another type of heterogeneity is environmental. So patients live in different environments, they live in different culture groups, they may do different activities. And so how their disease impacts them is really -- could be different, just as a factor of their environments.

And so all of these types need to be carefully considered when we're constructing both our clinical assessment measures, as well as endpoints, and providing context for interpreting those endpoints. And so, you know, how do we then approach this in a systematic way?

And so, I will again refer back to earlier guidance in this series. The very, very useful tool that we turn to time and again is our roadmap to patient-focused medical product development. And this was described in the third

guidance in the series on selecting or developing COA measures.

And it really starts -- I mean, the fundamental basis is starting with understanding the disease or condition. And so, this includes heterogeneity, it includes natural history, and the environment in which patients live, as well as -- very important -- the patient perspective. And so you will find that explicitly named in the roadmap.

And then from there, we also need to understand the medical product and what it does and how it's expected to impact the patient, and the specific patient population and subgroup that we're going to be targeting with this medical product. And this helps us then do what we call conceptualizing treatment benefit. And it really then drives the rest of the endpoint measure selection down the line.

So these are some of the fundamentals, and they really do apply throughout the guidance series. None of these guidances is really -- can be taken in isolation. And so I think that's a really important point.

And this final guidance on developing the endpoint, I really view as sort of keeping the end in mind. And I think that was mentioned earlier today. But it's so important to always think about the endpoint all throughout the development of a clinical outcome assessment.

And patient input is really critical. I mean, we've spoken with patients, for example, with rare, very pruritic diseases, and have asked them, what's the most important thing to you? Is it the severity of your itching? It is the frequency? Is it episodic or chronic or continuous? And what time of day is most impacted with the itching. Is it during the night or during the day? And so all of these factors are where patients can play a key role and influence how we define the endpoint.

And so, I'll just stop there.

LILI GARRARD: Thank you Elektra, for bringing that perspective through the other guidance. So we'll go with Lisa.

LISA KAMMERMAN: What I want to talk about is the rate of progression in different

conditions and how that affects the definition of a COA-endpoint. I'm also going to put it into the framework of the potential analysis and interpretation issues.

As you can imagine, I could talk a really long time on this topic. So I've distilled my short comments into a few brief topics and highlights.

The endpoint, in my opinion, the most important issues is the rate of decline. You can have a disease that declines slowly over time, has a moderate decline, or a really fast decline. And this has to be tied into the research questions question from earlier today that we discussed, and the scientific question more specifically.

So is the research or scientific question to show that your treatment in comparison to control is going to slow the decline, or prevent the declines, or improve the decline? And all those issues have an important consequence for defining the endpoint.

And finally, especially in rare diseases, there's always the concern of having a

mixture of patient population and different disease manifestations of the same condition.

This all translates into the length of the study and how long you have to conduct the study to show a treatment difference and durability. So a common type of analysis might be slopes. Is the rate of decline different? Is the change from baseline different, which would be a different analysis? And this circles back into the selection of the instrument you're going to use in the first place.

So you can imagine in a slowly progressing condition maybe very, very mild Alzheimer's, where you're using an endpoint -- (indiscernible) boxes -- I know that's in the news right now -- that declines perhaps slowly over time in a very, very mild population, where you're going to need a really long study to perhaps just show a small difference.

So in that case, even though you're probably interested in slope, it's critical to be able to select a tool that is going to have little noise, and to be able to detect a change over time. Because

at the end of the day, if there's no difference -- and this goes to the interpretation question -- is there's no difference because the instrument you chose was too coarse? Or was there really no difference? So all of these points really interact with each other.

The other possibility is change in functional status, and there are different types of functional status outcomes that can be defined. And how do you measure that? Especially then when you have a difference of functional status, for example, that baseline, is it an improvement to a particular category, or is it alleviation of all symptoms? But that's another possibility for analysis.

For longer studies, there's also the concern of missing data. So if you have -- and also, if the study -- if the disease is progressing slowly, maybe you can alleviate that by having fewer observations over time, as opposed to a shorter study. Maybe FEV1, which is performance based, obviously, and the six-minute walk test. So there you'd have a shorter study, but your missing data may not be as much as a problem.

Regard to mixture of functions at baseline, one way around that is to stratify your randomization. So if you stratify with high function versus lower function, the assumption is that the difference in treatments over time will be the same. So at the end, you can do a stratified analysis, which in effect averages over the strata. That might be one possibility in designing and analyzing such studies.

In the next question that we're going to address, I'll come back and talk a little bit more about rare diseases. I think they need more discussion. Thanks.

LILI GARRARD: Thank you, Lisa. And David?

DAVID REASNER: Yes, thank you. I'm going to make a few comments about baseline disease severity. And while Imbria Pharma does have a rare program, I'm going to focus on larger patient populations, because there are significant sources of heterogeneity, even in these more common indications. And if you think about diseases like seasonal allergic rhinitis, where time of year has an influence in terms

of the severity of symptoms, and weather impacts the severity of symptoms, so you have regional effects. You can get quite a lot of baseline heterogeneity. Or diseases like bipolar disorder, where it matters exactly what the presentation is at the time of initiation of therapy.

So in terms of baseline severity -- and some of these points may have been touched on at one level or the other in the other panels -- certain concepts may be endorsed at a particular level of severity. So that creates heterogeneity because you have certain sort of dead items, depending on the severity of the disease in a particular patient. They're not adding any information.

One thing to think about is how to summarize -- how to create the endpoint -- so that's an algorithm, a calculation -- and how to summarize the endpoint. Oftentimes we see percent change from baseline, which is an attempt to normalize the data. And it's an imperfect attempt, so the statistician on your project will use baseline as a covariant to take care of the remainder of the normalization.

But you might want to actually consider studying your endpoint on what I call the native scale, which is more intuitive, and you can label the axis according to how the COA was actually developed. Again, your statistician will take care of baseline by doing something like an ANCOVA model, treating baseline as a covariant, along with maybe other covariants that will remove baseline heterogeneity. But if you leave it on the native scale, it's much more intuitive.

The other thing I wanted to mention in passing at least is that you can't hear baseline heterogeneity by using extreme qualification criteria. It's often suggested, but in my experience what you're doing is driving a really high proportion of false positives in your selected subsample.

And in addition, you'll drive regression to the mean, which in and of itself is not an issue, until it begins to interact with something like a floor or a ceiling effect. 03:01:12 And now you've sort of run out of room to operate and it will create problems for you. So the solution is not

necessarily saying, you know, you have to have a 9 on a 10 point scale. Just FYI.

In terms of analysis and interpretation, at the group level you might want to consider the fact that there could be an interaction between baseline severity and your treatment effect. And that's part of what Lisa was talking about when she asked to consider stratification.

So you'll look to see whether those two strata can be combined, pooled or not. And if there are only quantitative differences, maybe a modest effect in one stratum and a strong effect in the other, you can pool those, because the inference is at a higher level. It's about all the patients in your trial. But if there's an interaction, then it's more complicated, and we look at things like heterogeneity of the slopes in the ANCOVA model, and you might need to summarize your data by looking at key percentiles, like the median. So those are things, again, the statisticians on your team should be looking at, whether it's part of your primary analysis or something that's in an appendix.

And in terms of analysis and interpretation at the subject level, I think John Scott mentioned this earlier. The differences that are meaningful and the differences you observe may depend on where the patient starts at baseline. So that's an interaction. And that is hard to investigate in the short time that you have a development program moving forward. Because what happens in terms of (indiscernible) development, you're likely using your Phase 2 data to bootstrap into your pivotal program.

If you start looking at the subsamples, you're working with smaller and smaller groups of patients. And you might have a group of patients at a particular level of baseline severity that are really just outliers. And if you give them equal weight in making decisions about what a clinically meaningful improvement is, you'll probably not make a good decision. So again, the goal is often to have a single criterion, which may or may not be possible, but looking at splitting your samples into smaller and smaller groups of baseline patients and expecting to

have reliable judgment from what is always a modest dataset, I think is probably not realistic.

And then lastly, I just wanted to mention that it's important to make sure your COA is sensitive to remission and symptom-free days. And I think that was mentioned in the first panel as well. There might be patients for which it's appropriate that you improve their disease to the point where they wouldn't actually have symptoms, or they might actually remit. You'll have spontaneous remitter's, but you may have remission due to treatment.

And sometimes I think we focus on the more severe patients. And ultimately, we would like to cure diseases, so please develop a COA that's sensitive so that if you should cure a disease, you will recognize it. And then ultimately, that level of change and whether it's important or not, you know, could be the difference between a particular impact, a proximal or distal impact, an activity that's a small change numerically, but to the patient, it could be an incredibly important change.

So, you know, this is patient-focused

drug development, so that's a question that you can put to patients. Thank you very much.

LILI GARRARD: Thank you very much, David. And Patroula?

PATROULA SMPOKOU: Thanks, Lily. So we've heard a lot about kind of general concepts of heterogeneity. So I'll give you just a few points from a perspective of rare genetic diseases, which is what we handle in our division.

So, you know, we talk about heterogeneity a lot, and of course, as we previously thought and discussed, it could be any type of heterogeneity. I think what we're talking about today is mostly for a typical heterogeneity in terms of symptoms at baseline, or over the duration of a trial.

When thinking about heterogeneity in rare genetic diseases, I think it's important to think about why we have heterogeneity. And of course, there's hypotheses and assumptions and some scientific (indiscernible) rationales, one being that genetic diseases affect individuals differently. There are many reasons for that. One simple reason may be

when you have a single enzyme defect, for example, phenylketonuria or some of the other endpoint errors, depending on what level of residual activity you might have, an individual may actually manifest different manifestations at different ages and also severity. So that sort of contributes to heterogeneity. The actual genotype contributes to heterogeneity, depending on how the actual genetic change affects the protein that's being made.

So those are some of the known factors. But there's many more that are unknown in terms of what causes this variability. And then on top of that, you deal with small patient populations, by definition, in rare diseases. So there's very limited opportunities to conduct really robust studies of untreated disease to really try to get a handle of why people behave differently or have different manifestations, and why that is.

So that brings me to the point of what we talked -- we always talk about natural history, so what this means, simply put, is what happens to a patient's disease when they are untreated.

And so, in rare diseases I think we rarely, if ever, have the luxury of actually knowing what the untreated disease looks like in patients. And of course, that's because we have a small population, studies are not doable, feasible, many times. There are small patient numbers, and on top of that, there's many patients who go undiagnosed for a long period of time, because those are rare diseases. And so you have even less opportunity to study this if the patient is actually not diagnosed.

The point to make that's very critical for the course of untreated disease is it gives you very critical information in terms of, of course, what endpoints to use, what instruments to use to assess those endpoints, but also what population would be appropriate for a given endpoint or clinical outcome assessment.

And the reason for that is because there is this heterogeneity, for instance, that I described, we see that, for example, children with a genetic disease have completely different symptoms than the adults with the disease, and they may have

completely different severities, and also disease trajectory.

And so if that's not known ahead of time, then it becomes extremely difficult to actually choose an endpoint that would be meaningful to the patient and actually specific to the disease.

The second point is we discuss a lot about patient input. And I think that's very much more important in the rare diseases, because we just don't have a lot of the knowledge of untreated disease. A lot of times, we don't really know what matters to patients because it's not documented anywhere. So, you know, how can you know what bothers them the most, what's most meaningful for them to impact with a potential medical product or treatment.

And I think Jessica Lee earlier mentioned that in our division we have kind of a history of engaging in listening sessions. So that's patient listening sessions. We have found that really useful, really important. Some of those that we've recently had was with patient groups with (indiscernible) disease. We saw a three percent

(indiscernible) adrenoleukodystrophy, phenylketonuria, and some others.

And so these are sessions where we listen. We pose questions. We listen to patients with, you know -- and we're talking basic questions. What's important to you, what's the most bothersome symptom, what would be really meaningful for you to see a change in? We get very different answers, depending on the ages of the patient. Of course, if you have a child is affected, you would ask their caregiver or parent, especially when you have a very serious or neurodegenerative disease such as (indiscernible) syndrome.

If you have adults, for example, with Fabry disease, the adult would very much be able to tell you what's most bothersome to them, and that could be their neuropathic pain or it could be their cardiac disease and exercise intolerance, just as a few examples.

Other points to talk about would be what factors really may determine what clinical outcome assessments or endpoints to select on a trial.

I think one major factor, when it comes to rare diseases, is the actual trial design. So if it is feasible, and many times it is, to do controlled trials, randomized controlled trials, then you have some of the freedom of actually doing work ahead of time and this is a hard to hook it scientists that are doing clinical disease trials interesting especially when there from India selecting disease-specific fit for purpose critical outcome assessments.

In certain diseases, though, where it may not be feasible or ethical, or other consideration of actually doing a placebo-controlled trial, you may be quite limited. So examples of that would be a single (indiscernible) trial where there is a historical non-current control group. So there, you're very limited to the selection of endpoints because many times you have to base that on what was actually done in the studies that you are planning to use. And that becomes really challenging. You have to really operationalize that, and we certainly have a few examples of that that was very, very challenging applications and were used to conduct.

In terms of other factors to consider, of course the baseline level of functioning, the baseline to the severity, and also the age of the patient become really important. And so it could be that within the same trial you have both pediatric and adult patients.

And then you may have -- you know, the pediatric patients have very different symptoms than the adults. An example of that is in Fabry disease, the adults have sometimes cardiomyopathy, they have renal failure. You don't see those in the kids. You really never do. The kids tend to have more gastrointestinal symptoms, or they have this heat intolerance.

So very different endpoint, but unless you have a handle on a natural history in relation to the patients, you wouldn't necessarily know. So how do you approach that?

I guess different approaches that we consider from the genetic standpoint or rare disease standpoint -- and of course, we discuss with sponsors and academic physicians -- are using potentially

multiple different endpoint, and not a single endpoint. I would say it's extremely rare that a single or primary endpoint would be able to really give you a good picture of what's happening with a patient with multisystemic disease. And on top of that, you don't have a good handle of the natural history, so how do you truly know if that one endpoint is truly representative, or even if you're going to see observable changes within a trial.

So I think the traditional paradigm of primary, secondary exploratory kind of really a lot of times doesn't truly apply to some of those rare diseases with multiple symptoms.

And of course, at the end of the day, from a regulatory perspective, we're looking for multiple lines of evidence in terms of evidence of effectiveness. And so having multiple different clinical outcome assessments, assessing disease-specific symptoms and being fit for purpose becomes really critical.

I guess another approach may be to have very liberal inclusion criteria in trials because you

have a very small population. And so restricting your population a lot of times doesn't make sense in a rare disease. But then it becomes very hard to assess for efficacy if you have this heterogeneity in the population. And that's where you try to become somewhat creative, like we always try to, and maybe select one of the -- one of those subpopulations are being your true primary efficacy population, and that could be the population which has the highest disease severity, for example, or a particular symptom. And I'm sure there's other methods that could be creative and novel in that aspect.

And then the final point would be this individualized endpoints and personalized endpoints. So, you know, we hear that a lot. We heard about responder indexes and such. I think from a perspective of the rare diseases I would deal with becomes even more challenging. So identifying --

LILI GARRARD: So, if I -- you know, we could table that for the next question?

PATROULA SMPOKOU: Yeah.

LILI GARRARD: Because we are going to

get into the individualized --

PATROULA SMPOKOU: Okay. So I'll save that for later, then.

LILI GARRARD: Yeah. Thank you, Patroula.

PATROULA SMPOKOU: (indiscernible)

LILI GARRARD: Those are great points. So, Tejashiri?

TEJASHRI PUROHIT-SHETH: Yes. So, good afternoon, everyone. I echo everything that has been set throughout the day, and I think Patroula covered a lot of the comments that I was intending to make as well.

And I think when you're thinking about designing a clinical outcome assessment for heterogeneous diseases, natural history is the key, if it is available. Certainly, it will help define, you heard, that it will help clarify what the disease manifestations are, what subgroups, whether they're children, adolescents, adults, you know, how these disorders manifest within these different age groups. You think about from the baseline disease severity a

clinical outcome assessment may be distinct for a pediatric population versus an adult population.

And the other thing I'd like to discuss a little bit is heterogeneity also presents in how the disease manifests from the waxing and waning perspective. So if you have disorders that wax and wane, such as multiple sclerosis or disorders that, you know -- David mentioned allergic rhinitis, seasonal preponderance of the symptoms, or environmental issues that raise the symptoms. And even in asthma, you can see that. You know, how do you design a clinical outcome assessment for a disorder that waxes and wanes?

So understanding the pattern of the disorder is very important. What causes it to wax and wane? Are there exacerbating factors that lead to this? Could these exacerbating factors (indiscernible) the concept of our (indiscernible) framework be considered in our current events?

But when you're looking at this, you also want to consider from a clinical outcome assessment, is the timing and the frequency of the

administration of the COA. If the pattern is they get symptoms, they may be free of symptoms for five, six months, and then maybe six months later, you know, that also impacts how long your study will be designed for as well.

So taking into a lot of these factors that you've already heard discussed are very important, and particularly important when you have a disease that has waxing and waning manifestations. And from a study design perspective, it is, especially in this context, very important to have some sort of concurrent control. It may be very challenging to utilize external or historical controls in this context.

LILI GARRARD: Thank you, Tejashri. And R.J., if you could help us wrap up this question, so we can move on to the individualized endpoints.

R.J. WIRTH: Yeah, it's hard, but --

[LAUGHTER]

All right. No more.

LILI GARRARD: (indiscernible)

R.J. WIRTH: No, I mean, there's a lot

of great discussion, and I think listening to all of it, it's obviously very hard. And it's a very complicated topic, and it's an issue that we all face, and what -- you know, Donna and me, when we were sort of prepping for this meeting, was that we never experience a single one of these. We usually end up experiencing a whole series of them at once.

And then you end up adding that complexity with missing data and restricted inclusion criteria. And before you know it, there's really just -- you know, you take something that's very hard and we make it harder.

And you know, unfortunately, I don't think there's a sort of -- even for any single one of these issues, there's no really straightforward single answer, because there's so many other sort of -- there's so much other variability outside of this. I mean, just the different types of diseases we study and the different types of endpoints that we have, the different types of COAs, the different types of analyses we can use.

So I think that what we can hope for --

I mean, I don't even think there's a complicated answer to this, you know. What I think we can hope for is for each case here -- to sort of reiterate what a lot of the people up here have said, you know, I think you need to really understand what the disease is. How does it manifest itself? How can it manifest itself in different ways? You know, what are the other external sort of threats to our studies, you know, missing data. Is it something that's seasonal? And can we come up with something that's our best shot for where we're at right now, right? I don't think there's sort of a silver bullet. I was talking to Scott earlier, that I don't think there is a silver bullet that's going to solve all of these problems.

So we need to sort of figure out in every given situation what are the biggest threats, what can we do, hopefully, a decent job controlling, and we just really think we need to understand what we're not getting right. You know, it's not going to be perfect. So can we figure out what we're missing so we can at least think about how that's going to impact how we interpret our results?

And that way, I mean, once we're upfront about it, then I think we can all live with that, right? Everybody say, I can't fix this, but I think this is how it would impact it. At least then we're being very transparent about it and we can take that into consideration when making decisions.

I won't go into a whole lot of detail on analyses, I think. You know, a lot of people mention them, and David and Lisa went into a bit more detail on specific analyses. But I do think we need to be open to look at new methodology, or at least new approaches. That's not really new, right?

But you know, if it's something that's cyclical and goes with the season, well maybe we should start thinking about using more nonlinear models, and we can model things using sine wave functions, we can take that into account. We can look at time variants and code variants of pollen levels, you know, from particular cities. AQll that stuff is sort of publicly available data that you could bring in as code variants to help control for some of that individual variability that -- from these external

threats.

So I think we just need to be creative, especially when we're dealing with these issues. And it gets even, I think, more complicated when we move to rare disease. I mean, when I think about trying to figure out how to control for these things, if we have enough data, we're not going to do a great job, but there's a lot of sort of tools at our disposal.

You start stripping away how much data we get, you know, now we have six people, and there's just not -- what we have at our disposal is just so much more limited that I think it just ratchets up the complexity that much more.

I love the idea about including more people and being more open, and then maybe having, you know, sort of a prespecified subset of the more patients that are more in line with what we typically do so we can see how that impacts the results. But any way that we can get more patients into the trials and get more data to bear on these questions, I think the better off everyone will be.

So my takeaway point is that it's hard.

And I think it's going to remain hard, and we just have to be really diligent and thoughtful, and if nothing more, really, really creative.

LILI GARRARD: All right. Thank you so much. Those are great discussion points. And you know, in the interest of time, I really want to hear your perspectives on the next question. And we don’t have a lot of time left, so if you could all give me a quick comment on what factors should be considered when constructing personalized or individualized endpoint for use in studies.

And I'm interested to hear from your perspective what the concept of personalized or individualized endpoints mean to you, and if you have any potential analysis or interpretation issues. But I think the main point here is trying to understand what we actually mean by personalized or individualized endpoints.

So I think the first sub-bullet point is something to focus on. So we'll start from Steve, our patient perspective.

STEVE ROBERDS: Thanks. I think this

might be oversimplifying, but I'd break it down into two things. One is what is the most bothersome, for lack of a better word, or most impactful symptom or symptoms to the patient who's participating, to the individual patients. And we realize that's why we're here. It may differ from patient to patient. And then the second is, what defines a meaningful improvement to that patient in that most bothersome symptom.

So a couple of examples, Elektra mentioned the migraine guidance. It was actually Dr. Billy Dunn who brought that to my attention a couple of years ago. And it's a simplistic way that I think about complex diseases like TSC. For migraine, it's pain, photophobia, (indiscernible) phobia and nausea. And so pain becomes the coprimary endpoint, and then each participant in the study picks which of the other three is their most bothersome. And so then you always have these two endpoints. It's just that the second one is different from person to person.

So again, it's a fairly simplistic approach, but I think it illustrates the potential

that we have to think about how to implement these personalized approaches.

But the next one is then what's most -- what's a measurement of improvement, what's the magnitude of meaningful improvement in those scores. And there was an example in the discussion document of the genetic eye disease and navigating -- you know, whether or not they could navigate this path under different lighting conditions. Really nice story. Really well-designed.

I was reading through this and then I got to the point where the meaningful improvement was two. What did I miss? What did I miss? Why is it two? So I don't understand why, and that might be something to add to the document, because maybe there's a story there. But why was it a score of two that was this improvement? Did patients say that's what was important and that the test is -- the study was with patients who --some have relatively mild -- it's all relative -- relatively mild impairment in the levels of light in which they can see, and others are very severely affected. Is it meaningful to improve

two steps for everybody, wherever they are on that scale? Or does it depend on where they start, you know, to the baseline points.

So I distill it down to those two things. So I think I'll wrap up my comments on this with maybe a pat on the back to Panel 1 this morning. One of the comments that I heard from a couple of people is having that patient as a partner in the design steps.

And I think one of the things that I've seen, being patient advocacy for the past eight years, is we talk very openly and clearly with FDA about what's important. And the FDA listens. You know, we've had a patient-focused drug development meeting, et cetera.

We talk with lots of companies. We have CDAs that -- can't tell you how many CDAs, not because it's confidential but because I don't remember how many. But we've got CDAs with lots of companies and we talk to them about what's important. But then the FDA and the companies go behind closed doors and decide what they can measure that's most important to

their patients, and then come out and tell us what the study endpoints will be.

So my point would be, how can we get the patients behind those doors? Can we bring them back, the patient-to-patient advocates, to be there at the table when the discussions are happening? And I think that could inform things like how do you decide what are the most bothersome symptoms, and how do you decide what magnitude of change is enough?

Thanks.

LILI GARRARD: Thank you so much, Steve. And you know, in the interest of time, and this is a topic that everyone has a lot to contribute, and we certainly do not run out of topics to talk about, right? But I do want to save some time for our audience Q and A, and then perhaps, you know, our panelists can interact with the audience and dive more into this particular topic.

So with that said, let's open up for question -- audience Q and A. And please directly yourself to the mics in the center of the aisle so we can have further discussions.

Do we have no questions?

R.J. WIRTH: Maybe it's not hard.

LILI GARRARD: Thank you. Go ahead.

MICHELLE WHITE: Do I have to turn this on, or is it...? Okay. Michelle White, Optum. I have a question with the most bothersome symptom. What happens when someone's most bothersome symptom changes throughout the duration of a three-year trial, perhaps as a result of the treatment they're receiving, some symptom gets worse? How does that affect your endpoint model?

LILI GARRARD: Elektra?

ELEKTRA PAPADOPOULOS: So, I'll take a stab at that and then, you know, others chime in, please. But you know, one of the things that we -- you know, we had a Duke Margolis meeting on this, a workshop, a few years ago on individualized endpoint. And one of the questions is, you know, do we frame this as most bothersome symptom? Because as you say, that can vary over time.

And so in cases where there's heterogeneity and symptoms among patients, our sort of

common advice that we commonly give in different situations may be to assess all of the symptoms at baseline and throughout the duration. Because as you noted, symptoms can change within a patient over time. And it's a particular problem in trials a very long duration and depending on the natural history.

And we also want to make sure that while some symptoms are improving, others are deteriorating at the same time. And so we really need to understand what is that clinically meaningful not only improvement, but also deterioration.

So those are just some ways. But clearly, I agree, with R.J. that it is hard.

R.J. WIRTH: That's why I'm here. Now, can I follow up on that, though? The most bothersome symptom -- I love the spirit of the idea, right? It's sort of when we're talking about patient-centered and sort of patient-focused, you really can't get much more than, just sort of what is it that bothers you, and let's measure that.

But from a measurement perspective, it's absolutely horrible. Because for one, we're

usually on a single item, which are horribly unreliable. And then two, what's most important can change over time. And then depending on how drugs or the treatment may impact different symptoms, you have people on different scales at potentially different severities, and where movement on the particular endpoint could mean different things over the different symptoms.

And it's -- I guess from a sort of psychometric standpoint, there's so many questions and potentials for error and messiness, that I think we're going to end up finding it harder to find treatments. And like I said, I love the spirit of it, but from a sort of statistical measurement perspective, I don't like it. So I like -- to sort of what Elektra said, recording all for them and tracking all of them. And if we can, think about a maybe more rigorous way to combine them in a meaningful way.

And if we are interested in the most bothersome symptom, at least have the other data there to fall back on if we are starting -- you know, if things do look a little messier than we were

expecting, making sure that we have that other data there to look at, and not just go, oh, wow, we really should have collected more. So I recommend the same thing. At least collect all of them.

LILI GARRARD: Lisa?

LISA KAMMERMAN: I think one thing we haven't discussed is the mechanism of action. So if we assume the mechanism of action of the product is the same for everyone and is of -- tempting to target the underlying problem, then having the most bothersome symptom, I think, is okay.

I think where you run into problems is where the mechanism of action may only be targeting a few of the symptoms. So in (indiscernible), there's a guidance on evaluating vasomotor systems associated with vulvar and agile atrophy. And in that particular condition, there are three coprimary endpoints, one of which is the most bothersome symptom. So women are receiving estrogen.

So in there, it's assumed that the estrogen has the same mechanism of action, and it's expressed -- the condition is expressed differently

with regards to symptoms. And that's a short-term study. I think it just gets more difficult when you have multifaceted symptoms in studies that run on for a long time. You don't really know what the treatment is doing.

LILI GARRARD: Thank you, Lisa. And then, Jean, you get our last comment for the session.

JEAN PATY: Hi. Jean Paty, with IQVIA. And for those of you that know me in the audience, I am going to make a comment and end with a question and do it in 60 seconds. And I challenge some of my colleagues to time me.

So a general comment, it's hard, but can we make it easier insofar as the whole idea of heterogeneity, personalized endpoints, to me, is coming up because in many situations, we are looking at rare diseases, fewer patients. And so this is the non-statistician approaching the technical panel.

At the end of the day, are we trying to understand is this patient better? Have we done something for them? So, to some degree, whether we need to go down the personalized endpoint route,

whether we need to think about most important symptom, et cetera, it seems to me that previous qualitative and quantitative data can be brought to bear for us to be able to come up with some metric that helps us understand, did we do something for this patient.

And that, to me, is our challenge. Our challenge to take this complexity and simplify it down so that we can just say, yeah, they got better, or they didn't.

So, R.J., in 30 seconds, give me the technical answer to that.

R.J. WIRTH: Was there a question in there? I didn't -- so --

[LAUGHTER]

JEAN PATY: It was directed to you.

R.J. WIRTH: Yeah, I know. I sorry, I would have listened.

[LAUGHTER]

R.J. WIRTH: No, well, I think -- you made me think of a lot of things. I mean, one of the -- obviously, one of the easiest places -- easiest places to sort of think about personalized endpoints,

and that is in rare disease, because it's a little more manageable when you have a very small set of people you're staring at.

But I also think the idea of heterogeneity is harder for me to conceptualize within rare disease because you have to have variability to have sort of heterogeneity. And when you only have a very small number of people, you don't really know, because you don't really have enough variability to know whether or not there's differences there. It's just you have like five random draws a huge population, they're not going to look the same. But is that indicative of the entire population?

So, I know. It's hard.

LILI GARRARD: Well, I suggest that you to get together during our breaktime and continue this discussion. You can have more than 30 seconds.

[LAUGHTER]

(Break)

EBONY DASHIELLE-AJE: Hello, everybody. It's now 2:15, so we're going to get started with Session No. IV. If you can make your way back to your

seats, that would be appreciated. Good afternoon, everyone. Thanks for coming back after break in a timely manner. I know we're toward the end of the day, so I also thank you for staying for our wonderful panel discussion.

Right now, we're going to begin Session IV, entitled, "Pulling it all Together -- An Example Across Guidances." Before we get started, I would like to ask my esteemed panelists to introduce themselves with your name and affiliation.

BILL BYROM: Perfect timing. Bill Byrom, I'm at Signant Health. We're an ePRO and eCOA provider.

MICHELLE CAMPBELL: Michelle Campbell from the Office of Neuroscience out of the Office of New Drug, CDER.

ANDREA CORAVOS: Andie Coravos. I'm the co-founder and CEO of Elektra Labs, which works and helps collect digital endpoints in clinical trials.

MATTHEW DIAMOND: I'm Matthew Diamond. I'm a physical medicine and rehabilitation physician,

and I'm the Medical Officer in the Division of Digital Health at the Center for Devices and Radiologic Health here at FDA.

MARK FRASIER: Good afternoon, I'm Mark Frasier. I lead the research team at the Michael J. Fox Foundation. We're focused on a cure and new treatments for Parkinson's disease.

ABIGAIL LUO: I'm Abigail Luo. I'm from the Office of Biostatistics and Epidemiology from the Center of Biologics.

ANDREW POTTER: Andrew Potter, Office of Biostatistics in CDER.

DIANE STEPHENSON: Hello, I'm Diane Stephenson. I'm the Executive Director of the Critical Path for Parkinson's Consortium, one of 15 consortia of Critical Path Institute.

EBONY DASHIELLE-AJE: Thank you all so much. So today's Session IV is a little bit different from the previous sessions because we're going to be discussing a working example. Based on earlier public discussions for this guidance series, we both heard from you and we've listened to you. We understand the

tremendous value for providing stakeholders with pragmatic illustrations that encourage deep thought about how concepts in each of the Guidances can be effectively understood and applies in practice.

And as a result of what we've heard from the public, we've decided to develop a working example that can be used to demonstrate the guiding principles of COA and COA endpoint development.

And our primary goal today is to begin exploring how to frame an example like the one that we're going to talk about in a little bit regarding DHTs to help stakeholders understand how to get from measurement concept to an endpoint. And DHTs provide an appropriate backdrop for this exercise because considerations for DHT implementation in clinical trials are relatively the same as the considerations for any COA. DHTs also offer an opportunity to consider the current topic of incorporating and implementing technology in clinical trials, and also what to do when you're exploring how to capture patient experience data, for instance, outside of clinic in a patient's daily life.

So during today's session, I'll first outline the brief guiding principles that I just mentioned that are applicable for COA development and implemental in general, and outline specific considerations within the proposed example of DHTs. Then I'll engage my esteemed panelists in a rich discussion as they reflect on the guiding principles and provide recommendations on what practical information we should include in a working example using DHTs as the backdrop. They will also discuss how these principles can be fleshed out so they can be broadly applicable to COAs in general.

And then lastly, but not least, we'll open up discussion to the audience for you all to provide input on the discussion topics. And I'd like to note, though, that while we're specifically discussing a DHT example, we want to limit discussion to the generation of COA data in development of COA endpoints using a DHT. Discussion surrounding details about DHT regulation and technical validations of DHTs will be out of scope for this discussion.

So first, I'd like to level set a

little bit and talk about why DHTs are appropriate. First, I'd like to make sure we're on the same page regarding some terminology. Digital health technologies are technologies that use computing platforms, connectivity, software and/or sensors for healthcare and related uses. DHTs span a wide range of uses, from applications in general wellness to applications as a medical device.

DHTs are also used as companion diagnostics, companion therapeutics or adjuncts to other medical products, devices, drugs and biologics, and they may also be used to develop or study medical products.

And when evaluating the utility of technology-derived study endpoints in clinical trials, there are a number of factors that we consider. However, at the center of it all, as we see with all COAs, we're most concerned with how to translate the data generated through DHTs into things that are meaningful and how to determine what would be the most appropriate technology-derived endpoints for a clinical trial.

DHTs include, but are not limited to, wearable, implantable or ingestible sensors, environment sensors placed in a subject's home, software applications, other general purposes hardware and specialized hardware.

DHTs can be used to assess existing endpoints or novel endpoints, and they may be used to collect data remotely. An advantage of DHTs is that they can capture data both actively and passively. Passive data capture could include those generated through accelerometers, cardiac rhythm measurement throughout the day, or actively through measurement during task performance or through patient responses as captured from a PRO.

Now, as you know, when we talk about COAs broadly, we often talk about the evidentiary considerations, and study endpoints used to support regulatory decision making and labeling claims must be based on well-defined and reliable assessments. And like other types of COA data when used in this context, data generated through DHTs need to be well-defined and reliable and should not be potentially

false or misleading when described in labeling.

Additionally, as with other technology-derived measurement tools, data and the subsequent records generated by DHTs need to be in compliance with FDA regulatory requirements for recordkeeping, maintenance and access.

So now I'm going to go into the guiding principles that we would like to reflect in a working example for this guidance series. First is the idea of concept measurement. We've covered this concept across Guidances 1 through 3. And within the context of DHTs, you know, we want to determine what are the most important concepts to measure by talking to patients and discussing these concepts with FDA review staff. This is the same as all other COAs.

But with DHTs, we also want to know for the concept or symptom identified, is the DHT an appropriate measurement approach to use to capture that data and measure that concept. If you've determined that a DHT is an appropriate measurement approach, then you have to assess if the DHT that you would like to use meets performance specifications, so

including accuracy, reliability and validity for the proposed intended use.

Then we also want to introduce a concept of usability testing. You all have heard this before; it's been covered in Guidances 1 and 3. You need to also with DHTs plan to conduct usability studies to ensure that the DHT is usable by patients in the proposed context of use without serious errors or problems.

Next, we have endpoint measurement, a concept that is being covered extensively in Guidance 4. So if you propose an endpoint using the DHT measurements, you need to capture the important concept that's been previously identified, and you also need to consider the statistical and measurement properties of this endpoint.

Lastly, after you have gone from concept to endpoint, you need to understand how are you going to deploy the DHT in a clinical trial; consider how to deploy and use the DHT in the study, including how patients will receive the DHT, how data will be collected from the DHT, and how clinical

operations will be adapted to incorporate into a trial.

So these are all common guiding principles that we have discussed throughout the Guidance series, from Guidance 1 through Guidance 4. And we have come up with a scenario within the context of Parkinson's disease that we think will be a good way to illustrate all of these guiding principles and then also bring it more specifically into the context of DHTs.

So the scenario that we're proposing is assessing gait in Parkinson's disease. So imagine -- ready? Put your imagination hat on. Based on a literature review, a sponsor asserts that gait, for instance, ability to walk distances, gait speed is important to assess in patients with Parkinson's disease. They're interested in exploring the use of a general-purpose consumer accelerometer to measure gait variability to support medical product development. They actually hope that the data can be used to demonstrate difference in gait variability between treatment arms in their clinical trial.

Existing methods to assess gait variability in clinical investigations are based on in-clinical performance outcome assessments. So the sponsor is wondering can a DHT capture data reflecting how patients function in their daily lives, should they be using this in their proposed trial. Everybody get that, processed it?

So now I'm going to engage our panelists We're going to talk through all of the different overarching guiding principles that I discussed in the slides, and then we're going to talk about things that would be useful to include in example with this particular scenario, highlighting DHTs as the backdrop.

So I'm going to have two questions for our panelists. The first one is: what additional details would be helpful to clearly illustrate the guiding principles as applied specifically to DHTs when the data is intended for use as an endpoint in clinical trials? The second question we'll talk about is: how well do the guiding principles illustrate considerations for any type of COA implementation in

trials, especially the importance of considering patient input and knowledge of the natural history of the disease when deciding on a target concept; for instance, gait variability.

So going to our first question. As I mentioned in the presentation, we have five guiding principles: first, concept measurement; second, tool selection; third, usability testing; fourth, endpoint measurement; and fifth, clinical study deployment.

So I first want to focus on the concept or the guiding principles of concept measurement. And as we know, you know, there are multiple best practices for ensuring that you're measuring the concepts that are clinically meaningful to patients in a trial. And we need to determine what are the most important concepts for patients by talking to patients and discussing these concepts with a review division as you're developing your endpoint strategy.

So first, I'd like to punt to my panelist, Diane Stephenson, to just talk about broadly the different types of considerations that we should highlight in an example related to Parkinson's disease

with regard to establishing one of the most important concepts to measure in patients.

DIANE STEPHENSON: Thank you, Ebony. So this theoretical example is not that theoretical for what we're doing in the Critical Path for Parkinson's Consortium. We've convened seven different companies with the Michael J. Fox Foundation and Parkinson's UK to engage early with the FDA. We had a critical path innovation meeting in May to really talk through these issues that are being raised today.

One key issue that we addressed upfront is the idea of target population. This was very well outlined with using Parkinson's as a case example in Guidance 1. And in the project we're leading, you know, really fascinating for me to listen through this last session on heterogeneity, Parkinson's disease is the fastest-growing neurologic disease, incredibly heterogeneous, both within and between patients, and different stages of the disease have different manifestations. The historical concept that Parkinson's was only affecting motor function is

completely changed now with the recognition that non-motor symptoms can be even more burdensome to patients.

So we pay close attention in our project to the targeted stages of the disease. We're collaborating to go to the agency early to seek feedback on a case pilot study called Watch PD that includes collection of data on a mobile app, as well as a watch, to look at features such as gait.

In this example, we aligned to only choose patients that were within two years of diagnosis of Parkinson's disease. This is a very important concept aligned with the stages of the disease that companies were planning their trials on. But if we had chose at different stages of the disease, it may have been a different concept.

So I think this topic we've heard over and over today about the target population is really key and that will really have a key aspect of importance as you start refining your analysis plan looking at something like gait.

EBONY DASHIELLE-AJE: Michelle, I'd

like to call on you to expand on the target population concept.

MICHELLE CAMPBELL: Well, thank you Ebony, and thank you Diane for that, and I am familiar with the work that Diane's consortium is working on. I think the message that I've been hearing recurrently today from the other panelists is really making sure we understand, when we are selecting a target population is, what is the natural history of the disease course, what is going on -- and that really helps us figure out what is that target population going for when we have to consider a drugs mechanism of action. So we want to make sure we're going to be applying that correct target population when we're studying that.

And I think the careful consideration in the example that Diane's highlighted is examining where do I think I can target the population right now in a disease with a heterogeneity. Often, people will just want to, when we're looking at considerations for clinical outcome assessments when we have a wide range of symptoms or a wide breadth of age span or disease

course, it can be a daunting task to figure out where to start. And so, really understanding where that disease course is and that natural history is a fundamental piece that really is important a piece when we talk about developing any type of clinical outcome assessments. And it all starts also back with making sure that we talk with our patients as well.

And so, in order for us to be able to understand that natural history, it absolutely starts with our patients because they are our experts in their disease. So I think that's why the population, target population, is really an underpinning backbone of when we're really thinking about how we're going to develop a COA, how can we translate this to an endpoint.

EBONY DASHIELLE-AJE: It's great that you guys are bringing up target population, and the key to making sure that you are including the right population, but also understanding the population before determining what concepts are most important to measure. We get a question from multiple stakeholders about, well, do we need to generate new data, or how

can we leverage existing data that's been generated regarding, you know, patient input in the area of what concepts are most important to measure.

So, Bill, I'd like you to speak about the importance of leveraging existing data and how things can be done to properly identify the most important concepts.

BILL BYROM: Yeah. Thanks, Ebony and hello everybody. Just building on what Michelle and Diane said about the importance of the patient and their voice in this process.

As I looked at the case example, it wasn't very clear really from the description of the literature review whether that literature was referring to perhaps clinician opinions or other sources of ideas, or whether actually that research was done with patients to really find out what was important to them.

And, you know, it's interesting because when I look at the activity monitoring literature -- and this is an example where we're considering using an accelerometer -- there aren't any examples that I

found in published studies where the endpoint that is being reported, or the set of endpoints that are being reported are defended in any way through initial evaluation with patients. Usually, it's a researcher judgment -- you know, I'll measure total steps per day or I'll measure, you know, the amount of time in vigorous activity -- and it's their concept that that's going to be important for the patients.

So it's refreshing to hear what, you know, Diane's describing in terms of the CPATH work, which is clearly going about this a very different way and the correct way in terms of involving the patient.

But I think, you know, what I'm interesting in really is just how we then do that. And, you know, going out to patients to understand, in this example, you know, what is meaningful to you as a patient in terms of your mobility. And, you know, it might be that a patient might say, as part of that research: Yeah, I'd like to walk, you know, without any motor function problems; I'd like to walk without freezing my gait; I'd like to be able to walk at an even pace for a number of minutes and I can't do that

at the moment, but that would make a big difference to me. And if that's the case, then measuring something like gait variability would sort of make sense.

And so, starting to develop a meaningful aspect of health, which is describing these important aspects of movement that the patients are describing; that's the first part. And so, we might then say, well, this reasonable that gait variability could be a meaningful aspect of health for this group of patients. And if that's the case, we can then develop a concept of interest in our studies to try and measure that.

And, you know, if gait variability is the meaningful aspect of health, what could the concept of interest be? Well, it might be something like, you know, the number of purposeful walking episodes that I'm able to perform without freezing, or it might be the stepping of all the time it takes between steps and having that, you know, constant, or the cadence, the stepping rate, and showing that that's a relatively constant thing; those might be reasonable measurements.

And so, as we start to think about then developing an endpoint around that, that's where we start to think, well, what's the best way to measure that. And it could well be that in this case an accelerometer is the best way to measure this. But it's only really when we've been through this process, starting with the patient voice, that we actually can then decide, well, what's the best measurement approach.

And as I started, again, thinking about this example, the gait variability, you know, there are different ways we could measure that. We could give somebody a diary or we could give an observer, or a carrier a diary for them to assess this for the patient, or we could put this accelerometer on the wrist and hope that we can measure this concept accurately.

And, you know, one of the things that's interesting about activity monitoring is that, you know, you collect all this data, but you rarely have context for the data. And so, you know, for example, if I'm looking at somebody walking and the variability

in their gait or maybe the freezing patterns in their gait, have they frozen or have they stopped moving because of a motor problem or did they simply stop moving because they were stopping to cross a road or they were doing something else.

And so, those are the sorts of things where this technology might not be ideal in this situation; whereas, a patient-reported outcome could actually be able to measure that even better. But I think those are the things we have to think about as we start to consider what is the best measurement approach for that endpoint.

EBONY DASHIELLE-AJE: Thank you so much, Bill. Now I'd like to call on Mark Frasier to just give us some highlights with regard to concept measurement from your personal experience with your work, just briefly highlighting things that you've encountered and best practices that should be highlighted in an example related to exploring what the best concepts are for measuring.

MARK FRASIER: Sure, yeah. So the Fox Foundation is supporting a study called Fox Insight,

which we're really excited about because it brings the patient voice into a study that is data centric and voice of the patient centric. It's an online study that can be -- that participants can take part of from the comfort of their own home; they just need a Wi-Fi connection.

And we have about 42,000 individuals enrolled in the study, so it's quite robust in terms of the data collection; 75 to 80 percent of those individuals are people with Parkinson's. And they enter information every three months, every quarter, and fill out validated questionnaires, as well as some more exploratory questionnaires that provides information about what they're experiencing as it relates to Parkinson's, what's bothersome, what's troublesome.

And what's been exciting to see is now we have this longitudinal data in some of the analyses that have been done. Particularly, I would highlight some by Dr. Ira Shoulson who is here in the room is that it's been reported what is bothersome, and that has led to concepts that have supported the initiative

that Diane mentioned in the critical path for Parkinson's. So we are marrying what's important to the patient, spoken by the patient in a data-driven way, with what to measure and how to measure it, so that's been really exciting.

EBONY DASHIELLE-AJE: Great. So next, we're going to move on to the concept of tool selection. So, you know, we discussed guiding principles for determining how to select the most appropriate DHT to measure the concept of interest. And, you know, you have to assess if the DHT meets performance specifications for the proposed intended use.

So for this question, I'd like to start off with Matthew Diamond from FDA to just talk about, you know, re-emphasizing what is most important for determining whether a DHT is the most appropriate approach, and then what aspects of DHT selection need to be considered.

MATTHEW DIAMOND: Thanks, Ebony. And I think it's important to first just acknowledge that for this example, I think it is very appropriate to

use a digital health technology because it does serve to illustrate the guiding principles. And just to go back for a moment to the first guiding principle of concept measurement, right, really understanding what's really important to patients.

I think that digital health technologies can be really well suited for that step and for this example. As a rehabilitation physician, part of my role is to really ask patients, understand how they're functioning and what their goals are and, you know, very concretely. It might be that a patient wants to walk up the stairs to their apartment. And I think here too it's important to really ask patients and understand very concretely what's important to them.

When patients talk about wanting to feel or function a certain way, it's really about doing that within the context of their daily lives, and digital health technologies allow measurements to occur in their native environment. Similarly, when patients talk about wanting to feel a certain way, it is over time, and digital health technology allow the

collection of information in between those punctuated measurements that might occur with more traditional measurements.

I think it's really helpful when the measurements fit seamlessly into peoples' lives. And if someone is using a wearable and maybe a watch that they've already -- that they would wear anyway, it doesn't add any additional burden and it doesn't make them feel sick because of the measurement now that's happening in their life.

Using digital technologies allows us to reach patients that might have difficulty coming in and participating in more traditional assessments. And what I think is very exciting is the opportunity to collect novel measurements to really get at what's important to people. I think that if you look at all the other guiding principles here for tools selection, in which I know is, Ebony, what you asked about. Using the digital health technology in this example allows us to go through and really evaluate all the different choices, because when it comes to digital health technologies, there are many. You talked about

it really spans a spectrum from both wellness products to medical devices.

In terms of usability testing, digital health technologies may raise novel questions about usability, especially when someone is using it at home. And for endpoint measurement, as we're talking about today, having data collected continuously by a digital health technology, you know, forces us to address some of the novel statistical questions involved in proposing an appropriate endpoint.

And in terms of study deployment; again, there are novel questions, but that also have applicability across clinical outcomes assessment most generally. So I think for this example, I think that digital health technologies, and specifically this one, is appropriate.

EBONY DASHIELLE-AJE: Andrew, can you speak to just with regard to the specifications for a DHT, not necessarily the detailed advice that we would normally give, but more so within the context of selecting something that can measure something accurately, the concept of interest accurately.

ANDREW POTTER: Yeah. So as Matthew had mentioned, DHTs have an advantage where you can measure, you know, continuously, so you have -- and that's a little bit different than some of the stuff we've had. But when we go to take -- go to create an endpoint, we're going to take probably take those measurements, combine them, so we want to have a tool that can measure those.

And the accuracy on our endpoint is going to depend on the accuracy of the measurement, maybe how frequently it's measured, and then is the person going to -- is the patient going to wear it. So if it's a -- you know, we may want to say we'll sacrifice, you know, so we want to consider these different things. So, for example, maybe we have a watch that a patient likes to wear; it's a little bit less accurate than maybe a hip worn DHT, but we can say, well, we get to measure more of their steps and that may be an advantage or more frequently and some of the tradeoffs will change.

And then we can also go back and look at specifications that the manufacturer provides for

the tool and say, okay, we have very accurate measurement here. But as we go through to create our endpoint, our measurement may increase before we do our statistics, or maybe the other happens.

EBONY DASHIELLE-AJE: Thank you so much, guys. Now moving on to the next guiding principle of usability testing. We know that we want to make sure that stakeholders plan to conduct usability studies to ensure that the DHT is usable by patients in the proposed context to be used without serious errors or problems.

So, Andy, can you speak about just practical things that we should be outlining in an example related to usability testing?

ANDREA CORAVOS: Sure. So I think when you think about -- this factors back into the tool testing, as you just heard about with other groups, so if you're, like, figuring out which tool you want to use. So when you're thinking about a tool, I would say that you first want to make sure that these tools have more benefit than they do risk. And you heard that there might be some tradeoffs between accuracy,

usability and other components. I would say there are four things to think about, and I'll walk through each of them briefly.

One is accuracy and thinking about what that actually means. Some of that will be covered by this guidance and not. And so, I think some instances when we think about where this guidance plays, we should really think about how that ties to other pieces and make sure that this guidance links to that. So, for example, if you're testing the accuracy of ePRO, that looks different from a sensor, and that would be something to consider.

That then factors into things like usability, so the different type of tool would have different type of usability. I would say that you can think about this section broadly, whether or not the tool is useful, and useful is a combination of usability plus utility. So utility is whether or not the product has the features that you need, and then usability is whether or not those features are easy to use. Those are two things that matter, and we'll talk about this later in some of the deployment.

There are other things to also think about so. So digital tools have a number of different types of risks than other things than we've ever had. Most, if not all, are connected to the internet. Anything that's connected to the internet, it's not if it gets hacked, it when. And so, how do tools deal with their cybersecurity, and FDA has issued a number of different guidances around thinking about that. And then also data rights.

So these tools are collecting a whole bunch of different pieces of data. And so, as you think about whether or not it's useful and usable, it's really important to make sure that people understand how and when their data are used. And usability is not just for the patients and participants who are using the tools, but also the data engineers who are incorporating those tools afterwards. So if you're selecting a tool that has APIs that are not really well documented or you can't ingest them into your dataset, then the tool might be useful for the patient, but not actually for the people who have to do the statistical analysis of that

thing.

So the four things are: accuracy centered around verification and validation; two, usefulness, which is a combination of utility and usability; three, cybersecurity considerations, in my opinion; and then four, around the data rights and management for it.

EBONY DASHIELLE-AJE: Thank you so much. Diane, can you speak a little bit about the qualitative nature of usability testing?

DIANE STEPHENSON: Certainly. The main misconception -- I've learned so much in leading this project over the past year -- is that it seems that there's a lot of confusion around what would be deemed appropriate for when you select a tool.

So as you know, there is various definitions and terminologies for FDA acceptance. But I've been told many times that people think that if you use an FDA cleared device or an approved device -- 510(k) cleared device -- then that automatically means that that device will be accepted as an endpoint, a digital endpoint in a trial, so that's incorrect.

As we've heard today many times that definitions of how to define what would be achieved with an endpoint are quite unique. And it's also through the formal qualification process that's been discussed by FDA, that it's not required that you have a cleared device in order to qualify a digital measure through the formal qualification process.

So these are just some really important grounding that we try to continue to remind, especially sponsors who are selecting digital tools for use primarily as exploratory endpoints, but their goal really is a digital endpoint in a trial.

EBONY DASHIELLE-AJE: So in the interest of time, did you want to say something?

MATTHEW DIAMOND: Yeah, just to provide -- thank you very much -- just a little bit of additional clarity there from the Centers for Devices and Radiological Health. Clearance of a device or FDA approval of a device is clearance for marketing as a device to be used for treatment, prevention, cure, or the mitigation of a condition or disease, and you generally would not require clearance or approval of a

device by CDRH for use in a clinical trial. As we're discussing here, the most important thing is that that product -- be it a device, a medical device or not -- is fit for the purpose in the trial.

EBONY DASHIELLE-AJE: Thank you so much for that clarification. In the interest of time, I just want to move on to endpoint measurement. And I'd like Abigail to speak about just different guiding principles related to endpoint derivation in analysis considerations.

ABIGAIL LUO: Thank you. So I think today, earlier today, we talked about an estimate of framework; basically, how you come up with a (indiscernible) question of interest and align your design counter analysis and interpretation of the clinical study so they can better align.

So I think it's very important you actually enable, very difficult but really necessary. And I think it's overdue interdisciplinary discussion at a desired stage. So I remember, I would like to tell -- so I remember when I first joined FDA as a statistician. So I was expecting the clinician would

just hand me an endpoint and a design, and then I would just calculate the simple science and do the analysis and check whether there was pre-specified and I would be done. So but (indiscernible) to be the case.

So I would like to give a couple of examples: so I remember a long time ago, I don't have the perfect memory of that situation, but it was an oncology indication. So our clinician told me that -- I read the endpoint was very convoluted. So basically, (indiscernible) told me that I would like to look and compare the duration of response among the patients, but not all the patients would respond to the treatment. But they said, what's really important to us is I want to know the difference in the median duration of response. And then when I look at the endpoint, I said I can tell you that both median would be the median, both median would be zero because less than 50 percent of the patients would respond to the treatment in either the treatment arm or the control arm.

So actually, that was -- in that

situation, that we had a lot of discussion between us, the statistician and our clinical colleagues, and we actually realized that actually you have two groups of patients in there: one group of patient would enter the study without the disease and you actually watch for when the disease recur. And the other group of patients, actually they would have the disease, you would actually watch whether they respond and then the duration of response. So it was much more complicated than what a single endpoint can capture within those two groups.

And that started my fascinating journey that I constantly talk to my clinical arm. So just back up a little bit. I'm from the Central for Biologics. And some people ask me, what disease area do you work on, and we actually work on anything under the sun that you can think of that a gene therapy or a cell therapy may be indicated for.

So I think the (indiscernible) was very important. And actually, in terms of developing the endpoint, I think we would benefit to have more earlier involvement of the statistician. Just as the

(indiscernible) said, sometimes you may have a clinical meaningful question you would like to answer, but it may not be easily quantifiable, so you may not be able to have a good estimate of what you want.

And then you probably would need to have some compromise; your primary endpoint would be something that's more easily quantifiable, but also clinically meaningful. But then you will also have rigorous collection of secondary endpoints and have supplementary analysis that will give you a vast answer to the clinical questions of interest.

EBONY DASHIELLE-AJE: Thank you so much. Now with regard to, you know, endpoint measurement within the context of the DHTs, there have been some things that I've heard before regarding DHTs uniquely having such high volume, high frequency data; and having so much data that's generated through these tools that, you know, it's necessary to pre-specify the window of time and the methodology used to calculate the concept response to treatment for each patient, describe how missing data will be handled, and then describe how the data will be aggregated to

generate a score.

So in terms of those types of concepts, we would like to also make sure that we are describing those sufficiently in the example so that people can consider them when they are trying to think about all of the guiding principles across the guidances.

Now we are running low on time, so the last concept of deployment in clinical trials. I'd like to punt that to Diane and have her just talk about what it means to use it in a study and considerations, your top two considerations that you would like to see highlighted in an example.

DIANE STEPHENSON: So thank you, Ebony. No surprise I'll say data sharing is key to success. And in the area of digital health technologies, it took quite some time for individual sponsors who are using these as exploratory tools in their trials to understand that it is going to be much more informative for them to share information and learn from one another.

But one of the themes of this meeting obviously is the voice of the patient. And so, we've

heard today over and over again about engaging patients throughout all the steps is so key. One of the many valuable lessons learned: we heard from both FDA and EMA from the critical path innovation meetings was how important it is to conduct exit interviews.

And we're so fortunate to be working with our partners at the Fox Foundation who've done such a great job at engaging patients in all of their work and validation of novel tools such as digital technologies. But the more we can engage the patients and hear of their experience, the better off we're all going to be.

In the traditional industry sponsors studies using such tools as exploratory endpoints, that information usually would not be shared. But I'll just say a call to action is we need to share this information with one another so that we reduce the burden and optimize the chance for success. These patients are incredibly inspired and want to help, and we've bene very excited to see how much these patients really want to adopt this technology in their lives.

EBONY DASHIELLE-AJE: Thank you so

much. And all of our panelists have expertly woven in COAs in general, as well as DHTs. So this last question I'm going to punt to Michelle. And just briefly in one minute, how well do you think the guiding principles illustrate considerations for any type of COA?

MICHELLE CAMPBELL: I would say that (indiscernible) should try to time me on this one, but I'm not going to put him to that test. I think actually I was really glad to see your slides earlier that really set out specifically those guiding principles and kind of referenced to people as a reminder where to find that information.

I think what we see here in this example and the combination of this meeting from the entire PFDD guided series is how all these guidances have overlaid and are interwoven together and should be used as an entire series when we're looking at how to incorporate the patient voice into clinical trials and ultimately into what is at endpoint.

And so, I think the guiding principles I've listed are the critical ones that are needed, and

we must ultimately start with what was in the first guidance, which was talking to patients and talking to the key stakeholders and opinion leaders and understanding a disease, and building upon that what their next guidance in terms of how do we take that information and kind of put it into an informative way, which will then ultimately lead us to, how are you going to select a COA and what's that appropriate COA. Is it here in this example, the digital health technology, or is it a patient-reported outcome or a clinician-report outcome?

And so, we need those foundational early guidances that we've talked about over a year or a year plus ago in this same room. Those guiding principles are the starting pieces to get to this point and the end of how do we take that information from -- we collect it from a COA or, again, this example here of digital health technology, and ultimately convert it into an endpoint.

So I think the ones that were selected are the appropriate ones. All of them reporting principles, but I think we need to remember to use

this entire series collectively together moving forward when we're examining how to incorporate that patient voice.

EBONY DASHIELLE-AJE: Thank you so much, and thank you to all my panelists for their insights. Here's a slide for helpful links. And now I'm going to open up the floor to audience Q&A. So if anyone has any questions or comments related to what you would like to see in a working example that's supposed to be cross-cutting across the guidances, feel free to come to the mic.

MAN 1: (indiscernible) from ICON. I've got a question specifically about the incorporation of multiple facets of a specific domain. In the example that you gave, you talked about gait used in the study, but there's other elements to that. So in gait, there's also going to be a difference of seasonality and how that's going to affect the gait and how it's being measured, so that would be a companion facet to that DHT that would need to be collected. Or in the case of a band where you're talking about measuring sleep; there's a difference

between measured sleep using a DHT and perceived sleep as collected with an ECO or a COA solution.

How does that tie into what needs to be collected and how that ties to what we were talking about earlier today in answering the right questions. I'd like, Bill, since you published on that, you'd have some --

EBONY DASHIELLE-AJE: Yeah, so I will punt it to Bill first.

BILL BYROM: Good question. And, you know what, I think I'm certainly an advocate that we need to do both. So, you know, using a digital health technology, if I'm going to use that term, if we're measuring an activity monitor or something like that, we still want to understand how the patient feels. And, you know, there's been some quite interesting examples in the literature.

So things like when we measure fatigue or we measure pain, quite often, a patient can show an improvement on a pain scale. And because of that improvement, they become a little more active. And when they become a little more active, they actually

feel a bit more pain again. And so, actually measuring the two together provides a really insightful picture as to what's going on; whereas, just one on its own may give you the wrong answer. So, you know, again, I feel it's so important to measure both.

And equally, actually what matters to the patient is actually how they feel. And so, if we stop asking them through patient-reported outcome measures, we're missing such a vital component.

EBONY DASHIELLE-AJE: Andrew, can you also contribute to that answer?

ANDREW POTTER: Yes. I want to agree with what Bill said about you still have to ask the patient. But also, for example, with sleep, you know, using a band with perceived sleep, but the same problem exists with polysomnography. What the sleep -- what they say -- the perceived sleep from a patient versus polysomnography sleep, that may also not agree. So you have patients who complain of they aren't sleeping, yet they go in and they have normal sleep time on a polysomnography and vice versa. So this

problem isn't anything new; it just may be that we have, instead of intermittent visits where every few months, we have to deal with this problem every day. So we've taken the same problem, just a lot more frequently.

EBONY DASHIELLE-AJE: So it sounds like the answer is yes, both. You know, you can generate complementary information from multiple sources, including that collected through a DHT and those collected through traditional measures or other COAs. Next.

SONYA EREMENCO: I'm Sonya Eremenco from the ePRO and eCOA consortium and consortia, two separate ones, at Critical Path Institute. And first, I'd like to thank the panel for an excellent discussion around these guiding principles. And I think, to answer the question, I do think that there is something potentially missing from these guiding principles in the context of these digital health technologies, which is feasibility. You talked about usability.

But as many of us, especially from the

ePRO world know, there's a difference between usability and feasibility, and it's something that's been explored in the CITI recommendations for deploying mobile technology in clinical trials that was produced, I think it was maybe about a year ago. So there's recommendations out there, but it really is looking at how well does it actually work in the trial setting. And it does involve things like talking to the sites who are working with the patients to administer the technology and how are things integrating together.

So I think that I just want to encourage you all to think about that, of incorporating the feasibility aspect as well. And I don't know if anyone had any comments and wanted to respond to that.

ANDREA CORAVOS: I know this might be untraditional, but the next speaker was the project manager for the CITI clinical technology, and she might be better suited than all of us. Is that -- can Jen answer that question?

EBONY DASHIELLE-AJE: Jen can answer

the question.

MICHELLE CAMPBELL: I think I would just say real quick before Jen answers, is that I think feasibility is an important thing. I'm sure Bill has some thoughts from the eCOA perspective of that aspect of how well can we implement when it gets down to the site. And so, it is an important consideration I think that applies to all COAs, you know. For example, if you're using a performance measure and it's not being standardized or across your trial sites, you're going to have data issues there too.

So I thank Sonya for bringing that point up, and it is an important consideration no matter what type of COA that you select to help support your clinical trial endpoint.

MATTHEW DIAMOND: Yeah. And just to add, and thanks for highlighting the feasibility question. And I think that if you really do tool selection well and the appropriate usability testing and think through the clinical study deployment in the underlying principles, I guess it's 2, 3 and 5, that

is really what you're wrestling with, that theme of feasibility across those principles.

EBONY DASHIELLE-AJE: Jen, would you like to add to that, and then also ask your question?

JENNIFER GOLDSACK: Sure. So I agree completely. I think the feasibility is a really important component to add to that really nice list of four considerations that Andy proposed. I also think it would be valuable to really be clear in the definition of feasibility study.

I think that, as we deploy more and more of these technologies, the time is being used really broadly and perhaps outside the original sort of definition that it had before we entered this era of digital. I think a lot of the vendors are using feasibility studies in multiple ways, and it's hard to actually identify which studies are going to give you the information you need, so coming up with some kind of taxonomy there would be really useful.

JENNIFER GOLDSACK: And now my question. My name is Jen Goldsack. I'm at the Digital Medicine Society now, having formerly been at

CITI. I wanted to make a comment. Diane, I was so happy when you brought up the issue of whether these DHTs need to be cleared devices. I've heard FDA twice personally, it may have ben said far more often, but twice say that these technology do not have to be medical devices and both times, it was in this room. It was during the launch of the CITI recommendations in the Summer of 2018, and then again today.

To bring some perspective from the technology side. Diane, you mentioned that there's a lot of confusion on the sponsor side about the need for these things to be medical devices. I think that that confusion is seen on the vendor side as well. And I would love to see it actually stated in the guidance, if that's appropriate, that these DHTs do not have to be medical devices.

And I think what that does is the next obvious question is, so how do we identify those tools that are suitable, and I thank the panel today. Thank you, guys, you did a great job of identifying what those characteristics might be. And I think if we can get away from this idea of being a cleared technology

being a way to rubberstamp the use of the sensor and the tool, it will really help advance the field.

DIANE STEPHENSON: Jen, I want to thank you. You've done a fantastic job at CITI and now in the Digital Medicine Society with so many -- addressing so many issues that everyone faced and also highlight the importance of standards, which your recent webinar tackled, so we all are very thankful to have you as part of our whole collaborative network.

MATTHEW DIAMOND: Yeah, and thanks very much, Jen, as well. And just to add one clarifying point. If the product that is intended to be used in a clinical study is one that is invasive and, you know, making measurements that are traditionally associated with a medical device, like a glucometer, it would be very surprising to see one used that is not a cleared medical device for that purpose. But if we're talking about accelerometers that, you know, would be used in a general-purpose environment, then it's really about, again, being fit for purpose. And being cleared in and of itself is not necessarily sufficient for being appropriate, certainly not.

EBONY DASHIELLE-AJE: Are there any other questions from the audience?

MICHELLE CAMPBELL: Can I just add one thing, Ebony?

EBONY DASHIELLE-AJE: Yes.

MICHELLE CAMPBELL: I think just taking Jen's point and the collective thing, and I guess it hasn't been said yet or maybe it has. But particularly as we emerge in this area with digital health as another option with digital health technologies, I do think this is somewhere where -- this is where the encouragement of that early conversation is important. And Diane's experience shows how that early communication was really important and critical to how they plan their study and next steps.

As we continue to learn about the capabilities and what kind of data and endpoint can be created from digital health technologies, in complementary with our other COAs, I think those conversations should happen early with your respective medical product center and division to make sure

there's agreement and understanding from all sides of really what are we trying to accomplish in the end. So I would just encourage that if you're considering that, making sure you've had some early conversations.

EBONY DASHIELLE-AJE: Thank you all. Just as a quick high-level wrap-up. It appears that, you know, from our panelists that the scenario that we're using in Parkinson's disease is one that's appropriate for highlighting all of the different guiding principles that span all of the Guidances 1 through 4.

And the guiding principles of concept measurement, tool selection, usability, testing endpoint measurement in clinical study deployment are all very important. But from our audience, we've heard that we're missing the key component of feasibility and, therefore, we should consider incorporating that into an example that we're going to be displaying.

And overall, we're thinking that all of these principles are not just applicable to DHTs, but across all COAs; and, therefore, it is useful for

highlighting the principles that we think span any type of COA that would be developed our used by our stakeholders.

So with that, I want to say thank you all for listening, and we are early a little bit. Right? I deliver, right? I overdeliver -- wait -- under promise, overdeliver, yes. So we have two extra minutes to chat, chat, chat, and then our panelists for the next session will come up. Thank you all.

(Break)

MEGHANA CHALASANI: Hi, all. I'm going to ask you all to please be seated so we can get started with our closing session for the day. My name is Meghana Chalasani and I work in CDER's office as a center director on the patient-focused development program staff and I have the honor of being the moderator for our closing and final session today.

The purpose of this session is to wrap up on what we've heard throughout the day and hear from our panelists on the key takeaways from the workshop. But then we're also going to broaden the discussion a little bit and ask our panelists and, of

course, the audience to reflect on the overall methodological PFDD guidance series that the FDA has been working on.

Before we get started with our moderated discussion, I'd like to ask each of my panelists to please introduce themselves. Marc, if you'd like to get us started.

MARC BOUTIN: Sure. good afternoon, everyone. My name is Marc Boutin. I'm the CEO of the National Health Council which is an organization created by patient groups for patient groups 99 years ago and has many, if not all, health stakeholders' representative membership.

STEPHEN COONS: Hi, I'm Stephen Coons. I am the program officer for clinical Critical Path Institutes, Clinical Outcome Assessment Program and I am an executive director of the Patient-Reported Outcome Consortium at Critical Path Institute.

KATARINA HALLING: Hello. I'm Katarina Halling and I am the head of patient-centered science within AstraZeneca.

TELBA IRONY: And good afternoon. I'm

Telba Irony. I'm the deputy director of the Office of Biostatistics and Epidemiology at the Center for Biologist.

LAURA LEE JOHNSON: Good afternoon. I'm Laura Lee Johnson. I'm a division director in the Office of Biostatistics in the Center for Drugs and I'm also the office Patient-Focused Drug Development liaison.

PANDU KULKARNI: Hello. I'm Pandu Kulkarni. I'm the chief analytics officer and vice president of biometrics and advanced analytics at Eli Lilly which is also including the Real-World Analytics Group. Typically, the statistics group don't include that, but that's why I wanted to call it out as a specialty.

MICHELLE TARVER: Good afternoon. I'm Michelle Tarver. I'm the director of patient science and engagement at the Center for Devices and Radiological Health.

MEGHANA CHALASANI: Great. Thank you all. So to get us started, let's start with hearing one to two of the most important messages that we've

heard related to the methodologies presented in the discussion document for this workshop and whether they can be reasonably and rigorously implemented in medical product development.

To get us started, Stephen, would you mind chiming in?

STEPHEN COONS: Sure. Well, the first thing I just want to say is that the utility of this fourth guidance is truly predicated on the quality and comprehensiveness of Guidance 3, because based on today's discussion and the discussion document for Guidance 4, Guidance 4 will assume that the sponsor has a fit for purpose COA ready to deploy in a treatment trial for the derivation of a COA-based endpoint.

And so that connection is critically important and hence, it's kind of difficult to fully address the adequacy of the contents of what were in the discussion document and our discussion today and will ultimately be in Guidance 4.

But both Elektra and Michelle mentioned that there are so many connections and

interdependencies between all four guidance documents or there will be and I think that is particularly true for Guidance 3 and 4. So hence, there will really need to be sufficiently clear and explicit cross referencing among the guidance documents since they are separate guidance documents.

And I was glad to know that at least the glossary of this Guidance Document 4 will be comprehensive. It will be an aggregation of all the glossaries from the previous three as well as the fourth.

And the only other point I'll make is that in the first panel, Kevin alluded to lack of information in the discussion document dealing with the heterogeneity in symptom and functional manifestations of disease within and among patients. And my assumption is the content of the panel three discussion will lead to a much more robust content in the resulting draft guidance around this issue and that will be incredibly important.

And Larisa in panel one and Steve in panel three both mentioned the heterogeneity in

meaningful within patient change in addition to the heterogeneity of symptoms and functional disease manifestations. And like (indiscernible) said earlier in his comment that R.J. didn't necessarily view as a question, we need to be creative and think about nontraditional approaches that may enable us to get closer to individualizing or personalizing end points.

And as R.J. correctly stated, it is hard. It's hard to do that, but it doesn't need to be considered impossible and particularly as more individualized gene and cell-based therapies are emerging. So I think we really need to put a lot more time and effort into thinking about dealing effectively with heterogeneity and getting to the point where we do have more patient-focused and personalized end points. Thank you.

MEGHANA CHALASANI: -- Stephen. I do want to emphasize one of the points that you made about the PF, Patient-Focused Drug Development glossary that encompasses the entire series of guidances and just teeing it up nicely. It is something that we envision being a living document and

so forth, so if you have an opportunity to review all of the terms and provide comments through the docket, that would be very helpful for us as well.

Same question to Katarina, just reflecting on the themes and takeaways.

KATARINA HALLING: So I think it kind of nicely builds on what you said, Stephen. I think the key things that I've heard today is a couple of questions related to Guidance 4 and where is it relevant to also include patients. I think that is something that will be extremely important to make that very explicit that it's not only in the beginning of the drug development and the planning that we do listen to patients and engage with patients.

It's actually an iterative process and I think we need to ensure that that is crystal clear throughout all the guidances, so it's kind of building on what you said, Stephen, about a little disconnect between some of the guidances.

And I think specific examples in Guidance 4 where I'd like to see more examples is, for example, in the clinical meaningful. In the

meaningful change section, there is a huge component where we take what we heard from patients in qualitative interviews and build that together with the statistical and analytical approaches.

I think in the patient burden section, I'd like to see a positive framing of what actually drives what patients think is a burden to them. There is text around what happens if it's too much burden, which I agree to, that's not good. But I think the key thing is that if we do include things that are important and relevant for patients, which we have heard from talking to them, I mean, that is one of the critical aspects of patient burden.

So I'd like to see a little more of that. I think that the estimand discussion was good. I've heard some say this is what we're doing. I think we're partly doing it, but we can be more crystal clear and be more systematic in how we do it.

Again, I'd like to see more examples of how patient actually also are alive in that conversation and I think that the estimand structure will be a good thing also to drive home what was

mentioned in the first session which is early.

So if we're really thoughtful in that process, it'll be easier to make the right decisions and start the right design of our endpoints already that early, and secondly, it will also facilitate the interdisciplinary collaboration which is critical to make this right. So those would be some of my comments.

MEGHANA CHALASANI: Thank you, Katarina. Pandu, would you like to add to this from the industry perspective as well?

PANDU KULKARNI: Yes, thank you. Let me first start by recognizing that in every single discussion we heard today, there was tremendous passion for patient-focused drug development. I think that's key, right. Everybody's aligned on that. Challenges will be there, but first of all, whether you want this or not is the question and I think looks like we all want this and we want this really bad and therefore question is, how will we make it happen and whether the guidance allows us to make that happen and what needs to be there.

So we have a great foundation to start with. So there are challenges. As we have been dealing with clinical end points for 50 years -- and these are hard end points -- and even there we have a lot of issues. So it's not an easy road and we heard that panel after panel; there are a lot of challenges. Even in the clinical end points recently we've been dealing with estimands and value of P value and so on, so forth.

So there is a lot of issues there we're dealing with. Now, we're introducing COA which has lot more variability, lot more unknowns and so there are really a lot more challenges and we need to be ready for those challenges. And I think the guidance touches on all of the aspects from having learnt from the clinical hard end points, so therefore guidance has done a really good job of looking at all of the aspects that needs to be done or dealt with and that's where I think the glossary is complete.

But what's not complete is the details of how to deal with it, and I don't know that that will ever be complete because this is an ever-evolving

arena and especial when you introduce digital into it and I think at lunch somebody was mentioning when she goes on a horse riding, her Fitbit thinks she is doing all kinds of exercise.

So I think those are things that we have to learn to deal with and they're going to be there and when these things are collecting data all day long, aberrations happen. How do we deal with those aberrations, I think, is a really key point. We don't know how to deal with them. You can't just take a look at the summary data. You've got to look at the data in a more fundamental way and we've got to develop new methodologies and I think guidance will have to provide some of those boundaries as to how we deal with it.

And the heterogeneity, in every single panel came up as to how there's so much more heterogeneity than you would have in the clinical end points and how do we deal with those heterogeneity and how do you make this homogeneous. If you try to make it homogenous, it becomes harder and therefore you have to learn to deal with heterogeneity.

So I think there are a lot of things that we have to deal with as a community and the key point that I think I took away from several of the panels was how important it is to collaborate now with different groups: psychometricians and clinicians and the clinical statisticians and people who have been doing the drug development for a long time.

We need to get them excited about these so that they can bring that experience together with patient-reported outcomes and COAs and that will make this whole process a lot better and smoother instead of people dealing with it all of a sudden fresh.

So my plea and urge will be to all of us to work together from the clinical experience to patient-reported outcomes experience. Bring it all together and deal with the uncertainty and the heterogeneity in a way that is manageable.

MEGHANA CHALASANI: Thanks, Pandu. Thank you for even the way you kicked it off by really recognizing and appreciating the patient-focused drug development or in this case medical product development, really, mission that we kind of have been

accomplishing through this guidance series.

Marc and I were just talking a little bit briefly earlier about how this was what we wanted when we were all at the table several years ago and how much we've really accomplished. One other theme that I heard through the day, a little bit throughout the different panels was some folks and some panelists were asking for a little bit more detail, some more technical details and other folks were thinking, we still need to understand the big picture a little bit more. We still need some high-level guidance and so forth.

And so really thinking and reflecting about who the audience is for this guidance and maybe getting some thoughts on, your thoughts on whether it's striking the right balance or how we could better strike the right balance, Marc, would you mind kicking us off?

MARC BOUTIN: Sure. Before I respond to the question, I just want to take a moment and say as a patient advocate that has been working in this space for more than 15 years and really sort of came

out of the concept of getting benefit/risk correct from a patient perspective, the patient community became really, really antagonized over the fact that critical decisions were being made for us without consulting or input from us.

And we pushed the FDA and then started to collaborate with the FDA and to be in an environment now where we have multiple guidances on device and drugs coming to the fourth one, I think we have to step back and recognize how far we've come. And so I'd love to get a round of applause for the FDA folks that have worked so hard on this for nearly a decade.

We in the patient community really appreciate the tremendous amount of work that's gone into this. From my perspective to your question, I think in this guidance we are getting at the right tone. We could certainly go more technical, but as you've heard from other speakers today, from a patient perspective, it's already a document that is very hard to understand.

And if you are what I call and average

patient, you're going to have a really, really hard time with it. So I think we need to think about how we present this to various stakeholders, how they understand it. My dream would be, we would have an infographic for each of the patient engagement guidances and we'd have a general infographic that would combine them together and explain how they work together.

Because we've had several comments already from this panel and previously that we need to think of them as a greater picture. And two things I'll share about that. One is, I don't want all of you researchers, scientists, statisticians developing the infographic. Want to be really clear about that.

Love you all to death. I do not want you to create those infographics. We have the contents, now we need to bring in experts that can do that and make them useful for what I would call real people.

And then second, we have all come together in this forum and I've heard a lot of comments where I listen to people and you get it. The

fundamental purpose of these guidances were to help us generate data and evidence that would inform us about representative samples and sub populations of people living and dying with disease. But I've also heard from some of you say, oh, we just need to do it the way we are. The core outcome assessment is something that's been new and ahead of the game for a long time. That's true, but at the end of the day, all of these concepts of engaging patients, codesigning, co-development, have to be front and center.

And just because you're in the business of developing core outcome assessments, doesn't mean that you're patient centered or getting it right. So it means we all have to bend. We all have to change if we're really going to get the impact we all want from this.

MEGHANA CHALASANI: Thanks, Marc. I'll turn to others on the panel to reflect on the theme about the right technical level and audience. Katarina, perhaps?

KATARINA HALLING: I just would like to start to say that I think you're absolutely right. I

think that we collaborate with different skills in this big endeavor and as you say, we're in a great place where we now are combining patient focused with good measurement properties and different ways of collecting that data including newer digital technologies. That's a huge step and we need great collaboration.

I think you're right. We should not explain this and be the ones that put those infographics together. I think there are people who can do that much better. I do agree with you that I think Gigi said it really nicely in the first session that it may be that it's not very accessible to patients and I'd love for this to come together and be understood by everybody which I know is one of the key things we set us up to do in the beginning.

And that will not only be good for patients. That'll be good for all of us in the collaboration to have a common understanding of actually what the little pieces are. Then, yes, we will need details for those of us who are doing this on a daily basis, but I do think we have most of the

details right, right now. If there's one thing I could wish for, it'd be, there's this summary of the guidance and recommendations of where we are today. I think we're recognizing we made a huge step forward to be where we are today, but also recognizing that we're continuing moving forward.

So just as I really like the guiding principles for digital health technologies, I'd love to see a framework for how do we continue to push for new innovations, how do we start tackle individualized measurement strategies. It is really hard. We need to continue to strive to make the complex easy and any frameworks that we could work together would be really helpful, I think.

MEGHANA CHALASANI: Thank you, Katarina. Pandu, would you like to chime in here as well?

PANDU KULKARNI: Yeah. Pretty much actually about the infographics, I think the infographics are extremely useful and they should be targeted. They shouldn't be just for patients. I think you also need it for other people. This

guidance is supposed to help everybody. Starts with the patients in mind and you do what the patients will get attracted to but you also need people who -- how to implement this thing.

And for them, you need to provide some details. Without those details, patients will get hurt in the end because it will take so long for the discussions to happen, you will never be able to develop anything fast enough. So I think you got to have, the infographics that I dream about, Marc, is if you are a patient, click on this. If you are a researcher, click on this. If you're a statistician, click on this.

And we have details for those roles and each of them can benefit from what they need to do. That way, you are not only having one in mind, all of us in mind because all of us have to do the work to get the patient benefitted. So I think this guidance is really good at the high level. It touches all of those, but it does lack those details.

Like I think the panelist said, I'm a psychometrician. I can see where it is going but I

don’t know what to do on a practice basis. Guidance says, come talk to us. You can go talk to them, but it will take very long time to have that discussion without certain guidance. So I think it would be phenomenal to develop that infographic for different groups. They can click on them and get some more details that they need.

You don’t need all the details, but you do need some more details for us to make this happen faster.

MEGHANA CHALASANI: Thanks, Pandu. I do want to turn to my FDA colleagues on the panel now and kind of reflect on what you may have heard so far already on the panel discussion but as well what we’ve heard throughout the day. Telba, if you’d like to get us started?

TELBA IRONY: Yeah. I will start with actually a point that’s very dear to me and is the discussion that we had this morning on the meaningful improvement or meaningful difference for a patient in any treatment, and what I heard combined with what I previously thought is that what’s clinically

meaningful, what's meaningful for the patient depends on the benefit/risk balance.

I think if you have a treatment with a high risk, a meaningful benefit will be higher than if you have a treatment with a lower risk. Of course, that depending on the context and depending on how much that benefit is needed and appreciated by the patient.

So we need to strike a benefit/risk balance to define what's clinically or patient meaningful, and also we need to consult the patients because what's clinically meaningful, what's meaningful for one patient, might not be meaningful for another one.

So how do we know what's meaningful for the patients? We talk about heterogeneity in disease responses and also in preferences. So I emphasize that's important to conduct patient preferences studies and obtain patient preference information, not only to determine what's clinically meaningful, what's a meaningful difference, but sometimes on choosing the end points and when you were talking about the

composite end points that have multiple components, even to give different weights to different components.

So patient preference information, it's very important and these studies are also very important. Now, going on the clinical outcome assessments, we all heard several times that they are not end points but they can be used to generate end points, and sometimes to generate several end points.

We also learned that we want end points that reflect patients' perspectives and also they are sufficiently robust to answer our clinical questions, our study questions and also to help us to make regulatory decisions and also to be included in labeling so the patients can make the decisions for themselves.

So it's important to align the end points with the study design and the scientific question that we want the study to answer and as Abigail mentioned in the previous session, sometimes one end point, one single end point, will not answer the questions. We might need more than one end point.

Which brings me to the discussion that we had before about the responder rates and the continuous end points. In that cases, in my experience, a responder rate, it's easier to interpret and it's probably what the patient wants to know. What will be my improvement and what's the chance that I'm going to experience that improvement?

For the statisticians, the averages, the continued end points and the (indiscernible) are more convenient and will be more powerful. In other words, we will require a smaller sample size for the study. So what's the solution for this dilemma? In my experience, it's to analyze both. We have one that will give us a better picture about benefit/risk per patient but also in respect to the overall patient population and sometimes for the public health.

So in this case, two end points will be much better than one even if they require more analysis. Also with respect to the heterogeneity of patients along time and among patients. I'm a statistician and I'm claiming the statisticians have a tool box to deal with the heterogeneity and with the

uncertainty that comes from the heterogeneity.

So what's in the tool box? Randomization. We'll fix a lot of these problems of reduced uncertainty. Stratification, modeling, and other statistical tricks. And I think that's what I would say. I also agree with you about the guidance and the accessibility of the guidance.

Someone also in one of the sessions gave a very good suggestion about making a patient summary for the guidance that maybe can be attached to the infographics and with all these things to patients will understand the content of the guidance but we actually need to be technically deep so that the professionals, the statisticians, the psychometricians will know what the FDA is expecting. And I'll finalize here.

MEGHANA CHALASANI: Thanks, Telba. Laura Lee?

LAURA LEE JOHNSON: Sure. So I'll be brief because you're going to hear a little bit more from me later as well, and I think we always learn a lot at these meetings and an important few points or

to clarify the links between and across the guidances, really the communication, and to echo Pandu, the collaboration.

I think inside of FDA, in many ways, we have worked on this very strongly but we're hoping that this also is more encouraging for a lot of our other partners as well in that area, that we need to write done some things that we would think were self-evident and several of our colleagues, like, yeah, you know we didn't say that because we figured of course. And we need to be more explicit.

And also having more examples. Everybody always wants more examples, so you'll hear my plea right around 5:00 to say, write them and send them to us. Put them in the docket because also you all's examples and what you're currently struggling with is going to be really relevant. We see what we see, but you all have a lot more experience there as well, so part of those examples, please send them and be willing to share them with all of us.

But it is the details and we also, this tension, the good news is we've been feeling this

tension for a while and trying to figure out what we were going to write and how we were going to write it. I do worry a little bit that if we separate things too much, do the roles and do the different groups then fully collaborate? Do they only read their portion and not the whole?

So I do think we need to figure out how we have both focused information but also that people really understand that whole. And so, I've said to my colleagues, I won't put all the details in an appendix because then people don't read the appendix. They focus on that key.

So how are we going to balance? And if you tell me I'm wrong, a thousand people tell me I'm wrong, maybe I'll change my mind. But I think these are some of the things that we continue to look at and we're really thankful for the feedback.

MEGHANA CHALASANI: Thank you, Laura Lee. And Michelle?

MICHELLE TARVER: So I think one of the things that we clearly heard is that it's important to put the question first, understand what are you trying

to address. And the reason why I bring that up as an important issue is that we're not only using clinical trials as the way of generating evidence to support product development.

We're looking at real world data sources that can then, therefore, inform us. So understanding what the question would be really can help with planning for registries as well, registry platforms. Understanding how is that end point going to be operational-wise can then inform how often should the general public that's providing information in the registry, how often should they be taking those clinical outcome assessments or performing those.

And that is important because you can't go back and retrofit it if it's going to be a comparator arm in a clinical trial in the future or used to inform the medical product.

The other theme that I think I heard was the least burdensome principle, and for our center, that's a central principle. We really do look and strive to make things the least burdensome, not only on companies but on patients as well as the

review team. We want to make sure we officially create solutions that can land in the public realm and help our patients protect and promote their health.

And so to that end, I think it's important to start with first things first, which is not just the question but the patients, having them at the table, having them part of the conversation, not by just having focus groups but actually having them as advisors sitting side by side and helping to inform how we're approaching these conceptualization, not only of the tools but of how they're going to be analyzed and the clinical investigation.

Our center has a draft guidance that we posted that talks about the importance of including patients' advisors in the clinical development and conduct.

And the last thing I'd like to mention is the importance of making the information interpretable. I think we heard that in many of the different panels and as a healthcare provider and ophthalmologist, when I have a conversation with a patient I'm not looking at their score on a

questionnaire, but I need to be able to translate that into language they understand so they can make an informed decision about their health care.

And so while it is important to have strong and robust measurement properties, it doesn't mean anything if it's not communicatable and so we want to make sure that we can communicate very clearly to the patient population and to healthcare providers who may not be as familiar with some of the language that we've been using in this room, be able to communicate that information so that people can make informed decisions. And I'll stop there.

MEGHANA CHALASANI: Thanks, Michelle. And so I'll look at my panel and see if, based on some of the comments that we've heard from the panel if that spurred any additional thoughts or sparked anything else that you'd like to share or add on before we transition. No? Okay.

I gave them very strict warnings about the need to, like, stay on time and be succinct, and so now they're not even going close to the mic.

So moving on to our second primary

question, which we kind of already started teeing up which is taking a step back from specifically guidance for the discussion document for this workshop and looking at the breadth of the entire guidance series.

We've already started hearing about the need to perhaps cross reference and link a little bit more between the individual documents and so forth, but just getting an understanding from you all, if there is a big picture that you're starting to see or that you see and how the pieces kind of fit together. Marc, would you like to reflect on that for us a little bit?

MARC BOUTIN: I think seven or eight years ago, we hosted a meeting and we actually held it here at the FDA with participation from the FDA to look at how we might move the concept of patient engagement along and we identified a number of barriers and the first and probably the most difficult was changing culture.

And if you think about how challenging that is, there's so many parties that have vested interest in the status quo. Even within your own

organizations and companies there are so many people with a vested interest in the status quo. And at the end of the day, changing culture is tough, but it really takes three things.

It's inspiration, information, and intimidation. And we have done a lot on the inspiration side. And it's not just the patient community. It's representatives of the regulator industry, researchers, all the folks here stepping up and explaining why we need to do better, why it needs to be different.

Well, one of the key reasons were so excited to partner with the FDA as it started to look towards guidance was that it would create the information, the how-to. And in all my years as a patient advocate, I've never heard industry say to FDA or any other regulator, give me more guidance, tell me what to do.

And we have set forth a whole lot of information. Is there more that can be done? Probably. Is there more clarification, can we get more technical? But I'll tell you, the roadmap is

there. The information is there. But most culture change, whether it's in society, in an organization, falls because we forget about the intimidation.

We're not hotwired to coerce people and nobody's asking the FDA to coerce you. But for all of us in our organizations, whether it's a nonprofit patient group or a company, we're going to have to coerce in order to change the culture and make this a reality. And that's going to be critical.

And for the people here who drank the Kool-Aid and are on board with this, you know when you go back to your organization, there are hundreds if not thousands of people who have no idea that this meetings' even taking place. And yet, you've still got to help move your organization to make this an integrated cultural reality.

Where FDA can help -- and again, I'm not asking FDA to coerce -- but this point about the integration of the guidances, this point about how and whether there's a clear understanding, I think for us in the room, the answer is yes.

But when I look at large companies,

this is going to be handed off into a very specific silo within a company and they're probably not going to look at the other guidances and they're not going to get the big picture. And if you do the COA work without the engagement from the very beginning, which has been said a number of times, you're going to get interesting results but it may not be relevant to the people you're trying to serve.

So anything we can do to drive home the fact that these guidances are connected, that there's a function that needs to be systematic, consistent, and integrated into the work that we do, is key. And I think we have to drive that message over and over and over again so that you can go back as leaders within your organizations and strategically move us to the next level where we put in those coercive components, where this gets written into all stakeholders' or all staff goals and objectives.

It becomes part of retention, recruitment, salaries, bonuses. That's going to be key to make this shift lasting and not have us revert back to the old culture where this becomes just a tick

the box and not truly meaningful.

MEGHANA CHALASANI: thank you, Marc. Stephen, would you like to...

STEPHEN COONS: Well, I do think there is a clear understanding of how they fit all together and I think -- and I addressed this to some extent earlier -- in theory, absolutely and I'm not being critical at all, but again, we haven't seen draft Guidance 3 yet, but in theory, we really start where we need to start in Guidance 1 and move to this point of having clinical outcome assessment tools that can then appropriately be deployed in clinical trials and analyzed and made into something that is interpretable and will be able to be put into labels that can help patients and clinicians make these important decisions regarding therapy.

So I do believe that there truly is an overall package here that is potentially very cohesive and comprehensive.

MEGHANA CHALASANI: Thanks, Stephen. Would anyone else on the panel like to comment on this? Pandu?

PANDU KULKARNI: I agree, Marc, the culture part I think is the critical part. I've been in the industry for 19 years and for the most of those years, PROs were an afterthought and they were just put in there just as other end point and nobody really thought through them very much and because there were not that many validated ones and very little chance of getting them on the label and so there was no motivation.

So my thought is still true motivation, inspiration rather than coercing. So the motivation and inspiration would come from having this guidance aligned with others to say, yes, if you did this, you would get it on the label. If you get it on the label, payer would pay for it. And if payer is paying for it, patients will benefit from it.

I think if we can get all of those elements aligned, I think industry is prime part of this. In every industry, if you look at every pharmaceutical company's website, they say patient is the number one. That's what we drive for. We go to work every day because we want to help the patients.

I think it is just that motivation to say, we can get this on the label, needs to be there, and I think this guidance is bringing that forward and then I think the payer groups also have to come along to say, if we have functional improvements, we will pay for this drug or device or whatever. I think it's those kind of motivation we need to drive and, of course, the culture change has to happen that we do need to think about these at the very beginning of the trial and even before that to say what is the patient journey and how can we take that into account in making up end points that are really useful and not an afterthought, and I think that would be very fundamental.

And having the discussion with the FDA, even to begin with to say, what can we be doing here to benefit the patient and can get it on the label, I think those discussions will really motivate people to do a lot more than they have done before.

MEGHANA CHALASANI: Thanks, Pandu and Marc. I've got a very excited panel. Very thought-provoking comments. Stephen and then I'll turn to

you, Katarina.

STEPHEN COONS: Well, I just wanted to recognize that we have use patient-reported outcome measures for years in terms of, there are certain symptoms, there are certain clinical outcomes that only the patient has been able to tell us for years. So I don't want to lose site of the fact that patient-reported outcomes have been very important. Pain, erectile dysfunction, all sorts of drugs have been approved based on patient-reported outcomes.

I don't think they were all necessarily as patient focused as we would like them to be, so I think that's what part of this whole effort is to make sure whatever we do in terms of clinical outcome assessments, whether they be clinician reported or patient reported or observer reported, that they are more patient focused and take into consideration what is meaningful to patients. So, just some clarity there.

MEGHANA CHALASANI: Thanks, Stephen. Katarina?

KATARINA HALLING: I do think that

we've started to see the culture change inside of companies as well and I do think that the effort with these guidance documents is helping us to put the patient at the heart of drug development so that will clearly help. That's also why it's so critical that we have a cohesive summary that we will all be able to use as a common framework.

And then just to comment on what you said with the labels, I also think that yes, labels is important. But I'd also encourage us to look for better ways of also incorporating the broader patient experience, even if that cannot make it into the label because we pre-specify one or two or three COAs, but there's a sea of information. I think we also can do better in representing the patient perspective more holistically as well, outside of the label.

MEGHANA CHALASANI: Thanks, Katarina. Michelle, did you want to comment?

MICHELLE TARVER: I was just going to say that there's a lot of development work that happens outside of companies, too. There's a lot of academic centers and other places where novel ideas

are being generated and having a nice layout of different ways to look, develop, and analyze clinical outcome assessments, I think, helps to facilitate some of those solutions that we're looking for for a lot of our health conditions.

I know that a lot of clinicians are doing research and when we can all sit at the table and understand what each other is talking about, I really do think that allow a generation of novel ideas and new methodologic approaches, then, can potentially be explored.

MEGHANA CHALASANI: Thank you, Michelle. I do want to touch upon whether there are any gaps, methodologically, approach-wise, for example. Are there any gaps, something that you thought you would see in this series, perhaps? Pandu? I see Stephen and Pando going.

PANDU KULKARNI: Go ahead.

STEPHEN COONS: Well, I do think that there are some of us here in the room that would like to see a little more on the analysis side and there's obviously a couple of examples in the appendix. And

again, this was mentioned earlier in terms of the first example, uses the estimand framework but it is more theoretical and doesn't have a complete sort of analysis plan laid out and it's not an actual approval; whereas, the second is a CBER approval, Luxturna.

And it is more concrete but it is not sort of laid out in the estimand framework and Kevin Weinfurt mentioned this this morning that he liked the example in Appendix 1 because it did have that framework and the example in Appendix 2 was less organized. But I think there could be a happy medium or both of those could be brought to the point where they really provide much more detail in terms of, particularly in the first one, how a concrete example could be provided that actually ended up with a label claim for a drug.

And then the other issue is just that there are many examples or several examples in terms of what FDA doesn't want to see like responder indexes and, essentially, percent change from baseline. But it would be nice to have just some more detail about

really what are the ones that FDA would expect to see and prefer to see, and I just think a little more detail there would be helpful.

MEGHANA CHALASANI: Thanks, Stephen. Pandu, did you still want to chime in?

PANDU KULKARNI: Yeah, so just one of the key things that would happen as we go into COA and then the digital is going to be tremendous amount of variation and missing data. Missing data. You look at the audience now. At the beginning of the trial was full. Now, it's about 50 percent missing. So the question is, how do I deal with that 50 percent missing and how do I motivate them to stick around in a clinical trial to the end of the trial? If not, how do I deal with it?

I think that part is really critical for us, and I think here in this example, it's going to be more an more critical, so I think we should head on, address some of that in the guidance because other ones, you can take into account and do what we have been doing in the clinical outcomes, but this, I think, is going to just be very, very tough to deal

with if we don't address it from the beginning.

So guidance should probably start to think about, how do I deal with the missing data, what kind of importation should I do, not do, and how do I deal with aberrations in the data if people are wearing Fitbits, it can go off for some reason wacko. What do I do with that data? There is no (indiscernible) on that data and therefore, what do I do with it?

I think those methodologies we haven't really developed them for big data, so I think that is something that we should figure out how to do that as a community and address some of the technical issues.

MEGHANA CHALASANI: Thanks, Pandu. Telba, did you want to --

TELBA IRONY: Yeah. I suggest it's part being participating in something that occurred to me as a gap when we're talking about end points making into the labeling, the guidance that is currently useful, the document emphasizes the hierarchy and hypothesis testing and always talking about making it to the labeling because you have to have several end

points and make it to hierarchy and it doesn't address a (indiscernible) framework in which you not necessarily testing hypothesis but you're talking about joint distributions of several end points.

So I think that will be helpful particularly when we are talking about strict control type one error which is very hard when we're talking about rare diseases in which the populations are small and the samples are small. So have to deal with more uncertainty in the strict hypothesis testing won't work as well. So that's one gap that we might want to address and the other one relates to what I said before, maybe we have to think about the another guidance on patient input and patient preferences, how to address benefit/risk, but maybe this is a big undertaking and we'll have to think more broadly about how to collect patient preferences to address benefit/risk determinations.

MEGHANA CHALASANI: Thanks, Telba. Laura Lee?

LAURA LEE JOHNSON: Sure. So I wanted to go back to this idea of thinking about, like, the

intercurrent events or missing data and things like that and part of what we need your feedback on is thinking about, like, some elements of missing data, like, this is bread and butter. It doesn't matter what your end point is. You got to deal with missing data in certain ways.

One part of what we tried to focus on in the discussion document is what are the weird parts of those COA-based end points? So like you mentioned, Pandu, like all right, now I have this accelerometer. What do I do with this as I'm trying to combine it and I'm trying to measure their walking or something like that? So these are the types of things that, you'll notice there are a lot of cross references in the document and we would like feedback because we can't kind of address everything-everything, but to the point we also don't want folks to kind of miss that they need to attend to something as well.

So just thinking about, again, that balance, but this is a big element of saying for a COA-specific end point, different than other types of end points, what do we really need to attend to,

because as you said, people haven't been thinking about it, so what are the reminders we need to give them, specifically because what they're trying to work from is that clinical outcome assessment?

MEGHANA CHALASANI: Thanks, Laura Lee. Katarina?

KATARINA HALLING: So just a quick comment on one of the things I think would be really helpful and that is, as much as we want, we would always like to move into our Phase 3 with perfectly fit for purpose COAs, but we all know that that's not always the case.

With the speed of drug development, in order to get new medicines to patients as quickly as possible, there are sometimes things that we can do in parallel to Phage 3 and I think it'd be great if we could have some commentary from the FDA on some of the acceptable things that you would be open to there, both statistically but then also in terms of confirming qualitatively in parallel to Phase 3 in order to have all the evidence when we need to look at the results together.

MEGHANA CHALASANI: Thank you. Does anyone else from the panel -- Stephen?

STEPHEN COONS: I just wanted to say that I was so glad to see a section -- and it's a placeholder for now -- on computerized adaptive testing because I think that is something, and I'm hoping that there are people in this room or people that are hearing this that will provide content to the docket that can be put into that section because I think it's incredibly important in the future because we are, through computerized adaptive testing, we would be able to have essentially short forms that are more personalized for the respondent.

And so I think that would help us in many ways and I'm glad to see through this document that the FDA is at least receptive to hearing more about the use of that in clinical trials.

MEGHANA CHALASANI: Thank you, Stephen. I do want to start asking folks if you have questions in the audience, please feel free to make your way to a microphone. I do want to leave ample time for folks. In the meantime, I do have one additional

question for our panelists. I mean, I have several, but start with one.

One of the more, beyond methodological aspects that the guidance series was tasked with kind of including was more procedural or process related which was getting at considerations for formatting and submitting the data as part of an application. And so I wanted to look to, primarily, our industry colleagues, the ones kind of putting applications together and submitting them, if those considerations are coming across clear in the guidance documents or if there's any gaps or any additional feedback that you have in that regard.

I don't know if, Katarina, you want to provide feedback or if Pandu wants to go first.

KATARINA HALLING: I think they're there. Again, I'd like to see even more clarity around where patient input actually influenced a decision. I think the patient experience table is very useful, but I think that's one of the things that we could probably improve more to be more clear on how patient focused we actually were throughout the

process.

MEGHANA CHALASANI: Thanks, Katarina. Pandu, anything to add? Oh, we have folks at the mic. Kim? Question?

KIM MCCLEARY: Hi, all. Kim McCleary. Marc and I haven't had the chance to tag team recently, so I'm going to take advantage of the opportunity to do that and just echo some of his thanks and congratulations to FDA for not only this series of documents but the tremendous progress that I think is very palpable in the room today.

Just a few observations of seven or eight years ago when we started down this road, you had CDER with the Patient-Focused Drug Development and CDRH with patient preference initiative and now look, all three product centers at the same table talking about these topics. Like, woohoo.

And thinking about kind of this higher level of what's the narrative, what's the story line, like, let's not lose sight of the fact now there are eight guidances, I think, total -- three from CDRH, and if you add the one about the guidance for

guidances to the four, that's a lot of information, as Marc said, a lot of direction and a clear signal to sponsors and industry and academia that you guys mean business as a whole agency.

This isn't some narrow little thing. What else have you written eight guidances on in the last seven or eight years as a package? And I think there's kind of a higher-level story that could be told about just this transformative process. I've heard Janet talk about it and Jeff talk about, it's changed the way you think about things. It's not only changing sort of SOPs and maps and your internal documentation, but it was really powerful today to kind of sit and hear the dialog and I had to go back to my agenda to figure out that comment that eight years ago would've only come out of a patient advocate's mouth is now coming from Michelle Campbell or it's coming from somebody from industry or it's coming from an academic.

Like, we've all put on different hats and switched roles and I think that's a powerful sign of change and the culture, and maybe it is a little

concentrated here in this room and a little more diffuse out in the other parts of the ecosystems we travel in, but it's meaningful and I think it is a sign of how to get the rest of our colleagues to move along with us.

So those are just come comments on maybe --

MEGHANA CHALASANI: Thank you.

KIM MCCLEARY: -- a little amplified of your (indiscernible).

MEGHANA CHALASANI: Thank you, Kim. Anyone want to -- no? Okay. We have one question in the middle.

CAROL MANSFIELD: This is more of a comment than a question. I'm Carol Mansfield from RTI Health Solutions and I wanted to echo what Telba said. The guidance documents cover a lot of ground, but one big hole is patient preference studies and I hope that that's on the list of things you're developing for the future.

MEGHANA CHALASANI: Thank you. R.J.?

R.J. WIRTH: Hello. First, I want to

echo what Stephen said about including computerized adaptive testing. I think it's great that it's a discussion and that it's in there and that we can provide more information, which leads to my question. It's been around for a long time. We know a lot about it, so with regards to what type of information would be most beneficial, if we can comment on that, to be submitted as part of the docket. Is there anything in particular that FDA might be looking for or is it just sort of, flood you with information?

MEGHANA CHALASANI: I'll turn to my panelists. Laura Lee?

LAURA LEE JOHNSON: I beg of you not to flood us with information. What I do want, though, information that can be copy and pasted is good, so think brevity, but usefulness. Not a thousand pages. But there's a difference between, you can have a lot of data and no information or a whole lot of information, right. What's key here and I think what the struggles are that we hear actually from industry more than just ourselves is how to implement in a clinical trial.

So as you're thinking about what to put in there, think about if you have a multiregional -- so this is an international, to use a less crazy terms -- you have an international trial. I got rare disease patients. I might have people, all sorts of different groups. I might have issues with connectivity. I might have all these other things.

How do you pull it off? So think about, and we hear concern about this, and also for most of you, like this is FDA guidance but I also want you all to think about what you've heard from other regulatory authorities or payers, et cetera, because as many people have mentioned, this information is going to move forward and the type of feedback we sometimes hear is, they know -- although, as you all mentioned, the point is not to stay where we've been but where we should move forward -- but they know they're going to be okay with this five-item short form.

So they're just going to go with that. And so to really understand and also when we're thinking about the real-world evidence, so now I've

got the tool but then we've got people that are collecting data in medical records and so it's part of a registry and a natural history study and now they want to use that as an external control or they're borrowing part of the information or they're trying to design their clinical trial, so how do we really implement it from a logistic standpoint, just like a lot of the DHT part, the digital health tools that we were talking about, but also think about the trial designs it could be used in and how we can also reuse data.

It's the reduce, reuse, recycle phenomenon. But it's in there because we don't want to say no, and too many people think the answer is just no and the answer's not just no. What we want to hear from you all is, what are the details, what are the struggles, what are the considerations, because a lot of what this document really is going to be in order for it to have a shelf life past the date of publication is to really say these are the processes and considerations.

So go forth and think and plan with

this in mind. Does that help?

R.J. WIRTH: A little bit. I mean, obviously, I will --

LAURA LEE JOHNSON: Yeah.

R.J. WIRTH: I won't send a thousand pages, half that maybe. All right, but it's curious and to hear about issues with connectivity and such, I mean, given that everything be app based now, that essentially ePRO, right, it's just either you get all the items or some of the items, but outside of that it's just ePRO. But I think in terms of just understanding sort of design and working that in, that will help sort of focus our submission.

LAURA LEE JOHNSON: And I'll also say, again back to the Guidance 3 and Guidance 4 part, a lot of the basis of computerized adaptive testing is item response theory and so thinking about all the different elements of that, if you want to write something that you think may be more Guidance 3 than Guidance 4, that's fine. We're open to that. Go ahead and send it in. Just tell us where you think it goes.

R.J. WIRTH: Thank you.

MEGHANA CHALASANI: Thanks, Laura Lee. Did anyone else from the panel want to add to that? Okay. We had one question up here.

KATY BENJAMIN: Hi, Katy Benjamin from AbbVie. There are a couple places where I'd really like to see more information. I was really hoping that there may be some additional guidance methodologically for the use and validation of other types of COAs besides PROs, especially (indiscernible).

I think that there are some real differences and challenges in validating these kinds of measures for use in clinical trials as patient-centered end points and how we go about proving that the kinds of things that are measured in these types of instruments actually are relevant and important to the patient. I don't think it's straightforward and I'd really like to see more work on that.

The other thing that I'd really like to see more on, because I think we're all struggling, is, as the panel has just acknowledged, the label is not

really sufficient to give patients a good idea as to the risks and benefits of a specific treatment. And we were talking about how else we can provide that information. Well, we now know how to collect all this PFDD stuff.

How do we actually present it to the regulators? How will it make it into the label or some other public format so that people can have this additional information?

MEGHANA CHALASANI: Thank you for that comment. Did anyone on the panel want to respond? Laura Lee?

LAURA LEE JOHNSON: So I do think a lot of the details you talked about in your first part of the comment really go more towards Guidance 2 and Guidance 3, but we note that and we'll try to make sure as we're getting those out the door -- and I can't remember if the docket for the draft of Guidance 2 is still open or not --

MEGHANA CHALASANI: It's open until the end of the year.

LAURA LEE JOHNSON: It's open until the

end of the year. This is why we love Meghana, because she remembers all these --

MEGHANA CHALASANI: One of the reasons.

LAURA LEE JOHNSON: But I would encourage you to put that in there because that, to look back through the draft of Guidance 2 to also see, especially if it's, like, do patients really find this important and are we gathering that for the non-PRO COAs, take a look and see kind of if that is tied in there well enough, and if not, give us your thoughts on how we could do it and put that in the docket, too. Thanks.

MEGHANA CHALASANI: And Marc?

MARC BOUTIN: Just very quickly to your last point, I don't know what the answer is but I think it is critical that we not only make sure that the information that is useful for other stakeholders -- payers, providers, patients -- makes its way into the public domain. One of the things that's so exciting about the work that has been done here at FDA is it has global ramifications, not just in drug development but in how we think about the health

ecosystem, how we deliver care.

And if we're ever going to get to the place where we really need to be, we need that information that is the foundation of all health interventions and it's the clinical research, and it needs to make it into the delivery system. And that doesn't all belong to FDA by any stretch of the imagination, but there's a key linkage to the entire ecosystem that is critical and it goes back to that intimidation piece.

If that information is there and it makes a difference to who is paying for care and who is delivering care and who is receiving care, and it's ultimately paid for, your companies are going to make sure that you do this in a systematic way from front to end in research. That's how intimidation gets in and shifts culture.

So I think this is a key element and I think we have to think carefully how we do it and I think it goes beyond the FDA.

MEGHANA CHALASANI: Thank you, Marc. We had one question up here.

MAN: Hello. This (indiscernible) from Gilead Medical Affairs Outcome Research Department. Actually I was trained 15 years ago as a Bell statistician, so here, I actually have two questions regarding in the current events here. And the first thing I think I would also acknowledge that several panelists already mentioned that these guidelines pretty much focus more on randomized clinical trials instead of real-world studies, right?

And I think after reading through these guidelines, I think one thing probably is currently missing is loss followup in real-world studies, like observational studies, especially for these patient-reported outcomes or COAs simply because of lack of motivation of the patients because in randomized control trials, it could be like a compulsory (indiscernible) driven procedure, but in such kind of a real world observational studies, it is not.

But it’s also sometimes not part of the daily, normal clinical practice so here, my question is to the panelists. It’s how we can mitigate such kind of thing in, like loss followup because of lack

of motivation from the patients and I mean, sometimes according to our experience, I mean the response rate from patients could be as low as, like, 10 percent, 20 percent from such kind of real-world observational studies. I mean, how low that the FDA or the panelists think that is it acceptable as such kind of COA studies? Thank you.

MEGHANA CHALASANI: Thank you. Laura Lee?

LAURA LEE JOHNSON: So, everything is situational dependent, so you're not going to see a line of how low can you go. But I do think we can do somewhat of a FAQ with some of us, the frequently asked questions or something like that. I would say that nothing is ever compulsory, really. One of the first trials I worked on, and we had a lot of COAs in there, so there was a huge burden.

This was an observational study and it's a huge burden on the people that were involved in it. And what we found, actually, was as people got better, they didn't have time to sit down and fill out all that stuff. They went back to work. So people,

like, oh, you're missing data must be the sicker people. And as we dug into it, I was like, no.

So this, though, gets back to also thinking about trial procedure, so regardless of what data you're missing, is thinking about talking and engaging all of these patients throughout the entire development of the trial, is this what they want to answer. Can they do it? How are we facilitating their being able to continue their participation?

So a lot of these issues, they're not like statistics. We're the dead end time. Like, now we don't have a choice. We've got to figure out how to fix it. The best time to do that is way earlier in that trial planning and having the consistent engagement and realizing also, like, hey, we're starting to see a problem.

Let's talk to our patients, do those exit interviews, do other things, and figure out, okay, we thought we planned well enough but now we see a problem. How can we redo it? And that's, in some ways, beyond the scope of this but in some ways it's tied. But thank you for your comments.

MEGHANA CHALASANI: I did speak with Mary Jo who's the moderator for the open public session after this and she gave me permission to go a few minutes over. So Pandu, did you want to add here? Oh, okay. Michelle, go ahead.

MICHELLE TARVER: So I think my point is going to be very short. I think it's important to measure what patients care about and we've talked about that already and you heard Telba already allude to ways that you can potentially approach that with patient preference information to inform those outcomes if you have to prioritize, then you can prioritize with what's most meaningful to them.

CDRH and CDER put out a guidance document on patient preference information. I encourage you to look at that if you're looking for some information on what that's about and how it could potentially be used in a regulatory context.

MEGHANA CHALASANI: Okay. Marc?

MARC BOUTIN: Just super quickly because I think you guys nailed that response. When you have a situation and you want to interpret it, you

always have to bring in the patient community or you're going to make mistakes. One of the classic examples I hear from the AI folks is, we can tell you when the child puts a wearable on the dog. That's great. Can you tell me why the child put it on the dog? No. But if you engage the children and their family caregivers, you can figure that out.

So bring us in to help you interpret those issues. We can make a big difference.

MEGHANA CHALASANI: Thank you. And Pandu, you wanted to jump in?

PANDU KULKARNI: Yeah, so I was going to just say the best missing data is no missing data, and so we got to start, as Laura Lee said, from the beginning itself. I think we don't spend enough time and research this on making sure that we get all of the data. Having said that, I actually, Marc, I think we could use patient advocacy groups and other groups that can help us with motivating the patient to stick around and provide the data so that it is useful.

Sometimes, it is not lack of motivation. It is because they got better and they

didn’t fill out and sometimes it is because they are just terribly feeling bad and therefore they leave. So there’s a reason why they’re leaving. We just don’t know many times why they’re leaving. And sometimes, it is just they’re losing, leaving out somehow.

So I think utilizing some of the patient advocacy groups, we have not, I think, typically done that but probably we should start doing that to see how we can motivate the patients, although the patients are very hard to reach throughout the globe when we are doing a global study, but we should do something about it.

MARC BOUTIN: Cocreate and we’ll help you motivate.

MEGHANA CHALASANI: Laura Lee?

LAURA LEE JOHNSON: And I know we have one more person, but I want to add into this, thinking back to that estimand discussion, which is to also go aback as you’re doing this because systematically collecting and following up and understanding where people ended up, you may be missing that particular

score but I have a whole lot of information and now you can turn that and say, okay, what's really that research question that we need to address. What really should the end point and the summary be.

And especially if you've done that prework, a lot of times, we know what could happen in advance, but did you actually work that into your estimand? Are you sure you can really address the right research question, the one you're going to end up with in the end? A lot of times, we can pre-think this and we should, but it's when that information isn't collected, it's not worked into there, because it takes money, it takes time. But that's something that if you want to do it robustly, sometimes we get that, we use it, and we can do that.

MEGHANA CHALASANI: Thanks, Laura Lee. And so for our last question, I'll just go to the middle right here.

MAN: Yes.

MEGHANA CHALASANI: And after that, we'll have to --

MAN: Yes. (indiscernible), Clinical

Survey Outcomes. Just like to follow up on Katarina's points in terms of the qualitative research and how that activity can support the drug development, not only in terms of meaningfulness but also to get a better understanding on the PRO responses and to challenge the product profile as well, because when you are in drug development, in the early stage of the drug development, there are great uncertainties around the product profile and the confidence intervals are very large and then, which PRO instruments should you then include.

And that is very often a challenge and are there additional benefits that are not seen on the product profile that you do not know. What you do when you're in this situation, I've been there several times, well, then you include some quality of life measure and treatment impact measure that has been validated already and then you hope that that will capture what you hope that the product actually will have a benefit.

But then, it's included in the early trials, these safety trials and they are conducted in

small samples, so the struggle continues. You actually sit there with PRO results once these result comes out and because of low sample size, you can only, at best, see trends in some of the scores. And for lack of better knowledge, what do you do? Well, then you include the same PROs in the following large clinical trials.

And then you may actually hear that patients come back and say, well, we have a struggle understanding the questions in the generic messages because we cannot relate to them. And then you actually then get the results. You may actually have (indiscernible) coming over and saying, oh, we're very excited. We have scoring, SF-36, moderate physical pain by 0.2. What does it mean?

I don't know. Is it less pain in the joints, less pain in the heads or stomach? Cannot tell. And then the product comes on the market and then only to realize that oh, there was an additional benefit that the standardized PRO instrument did not measure. Oh, we found out there was, the population that we thought would appreciate the product actually

turned out to be completely different and all because we didn't listen to the patients at some point.

And that's why to encourage you to put more emphasis on listening to the patients actively in the drug development, not only just using PROs. Thank you.

MEGHANA CHALASANI: Thank you for that comment. (indiscernible). I think with that, I think since I've taken enough of Mary Jo's time, I would like to wrap up our session. I want to thank all of my panelists. Thank you all so much for your participation and to the audience for being so engaging. And in case you have not heard, we have a public docket associated with this workshop and it's open until February 4th of 2020, so we encourage you to submit additional comments to the docket.

And with that, I'd like to invite Mary Jo.

MARY JO SALERNO: Good afternoon, everyone. As Meghana just stated, my name is Mary Jo Salerno and I work in CDER Office of Biostatistics. I have the honor of moderating the public comment

session and the tail end of this workshop. So for those of you that are not aware, the purpose of the public comment session is to allow an opportunity for those who have not had a chance to speak on issues that are not related to our main discussion topics of the workshop.

Please keep in mind we will not be responding to the comments, but they will be transcribed and be part of the public record. We'd like this to be a transparent process, so we encourage you to note any financial interests that you have related to your comment. If you do not have such interests, please state that for the record and if you prefer not to provide this information, you may still provide your comments.

We've collected sign-up before the meeting and during the break. We have four participants signed up and about 20 minutes for the sessions, so that'll be approximately four-minute time limit for each. We'll be keeping track of time. I'm not sure if it's really necessary, but -- two minutes? Okay, all right. I stand corrected. I was just

dividing 20 by four, so I guess this will just be a shorter comment period, so thank you for clarifying.

I'll ask you to wrap up, and as we've stated numerous time, the public docket is available for further comment and we encourage you to comment through the public docket. The public docket closes on February 4th, 2020.

So we're going to take our comments from the microphone in the middle aisle and if someone is not able to get to the middle aisle for mobility reasons or other reasons, we'll hand you the microphone. I'll run through the order of the commenters. Please note the name of the commenter before you and be prepared to line up at the microphone when he or she begins commenting.

So we have Andrew Trigg, Carrie Barnhart, Danielle Meyer, and David Reasner. So if Andrew Trigg can please go to the microphone?

ANDREW TRIGG: Yep, thanks. Yes, so I have no financial interests, first of all, but yeah as a bit of a disclaimer, I think I signed up to this this morning, but with hindsight it's something that's

been discussed quite a lot already today, so I'll keep it brief.

Yeah, so I'm a statistician working in consultancy and I think one kind of challenge we face and the challenge quite kind of linked to what's in the guidances is convincing pharmaceutical clients that it's okay to have kind of mean-based between group difference kind of end points in terms of your actual kind of hypothesis testing.

And I think based on what people have said today, it seems that there is agreement that it's okay to do that and your within patient stuff can just be, I suppose, supplementary, like you want to have it. But really, when looking at a guidance and when I think a lot of people we speak to read the guidance, you've got about a quarter of it is within patient meaningful change and it can seem, I think, that that is the only thing that the FDA are interested in looking at.

So I think a recommendation from really the statisticians in my team is to kind of make the discussion between the differences a bit more

prominent in the main body. I think at the moment, it's a bit kind of hidden in the appendix or in little areas and to tie in with that, really thinking about clinical relevance in terms of between group differences.

Again, there's so much focus on what is meaningful with patient change, but also tying that back and saying, ultimately, we will sometimes want to look at between group differences and means, but we also need a threshold for that as well. But we'll put that in the docket. Thanks.

MARY JO SALERNO: Thank you for your comment. Next, we have Carrie Barnhart. Carrie, are you here? I don't see anyone at the microphone. Okay. It looks like Carrie is not here. Danielle Meyer? Yeah, I don't think that Danielle is here, either. So we'll just have one more comment and that is David Reasner.

DAVID REASNER: Well, I'm laughing because actually my comment was covered in one of the more recent panel sessions, but not because there was a vacuum, but in response to the other comment, I'll

make a short comment, and I don't have any marketed products or affiliations other than I work at Imbria Pharma.

The comment, a number of panelists commented on this dichotomy between testing for mean differences and testing responder end points, and I actually think we muddied the waters today, so maybe it is worth making a comment. It makes sense to study mean difference if you're hypothesis testing and you need a primary end point. And I'm glad to see that that's represented in the discussion document because that's usually my recommendation to teams.

But that said, we do a lot of work to come up with a clinically meaningful improvement talking to patients, so we should use that in interpreting that mean difference. And in fact, in explaining these topics to folks, I've come to believe that the easiest thing to say is, there's no such thing as a clinically meaningful difference at the group level.

It only matters if an important number of patients get dragged over the threshold of an

important threshold clinically. So in that sense, put that end point wherever you want, but you won't understand your trial until you look at a range of response that's been endorsed by patients. So I guess that's the two cents.

It's really just one set of data. They aren't different data. In fact, there's pseudo specific. So it's two faces of the same coin and you can test the mean difference, but please look at a responder definition prior, preferably a declared a priori. Thanks.

MARY JO SALERNO: Thank you for your comment. Okay, so brief public comment session, I am going to now turn over the microphone to Laura Lee Johnson for our closing remarks of the day.

LAURA LEE JOHNSON: So thank you all for sticking around this late, I will say. And I want to give first a little bit on process because as we've mentioned many times, please send us your comments. This is what that page looks like and so for you or anyone who wants to provide comments either on the meeting itself, so what you're heard here today, on

the discussion document, send the details.

Please list things like the section of the document if that's what you're talking about, the line number so we can more quickly work to combine the comments and address them as we work on draft guidance. And if you think that your comment might be relevant to other guidances in the series, just let us know that, too, or other guidances in general.

Now, just to give you a little bit of feedback, when the draft guidance also publishes, there will be a docket for public comment there as well, so this is not the one and only time to give comment. There will be multiple opportunities to give comment. You can also look at our Patient-Focused Drug Development web page for more details on the process.

So I do want you all to remember we have questions for you. And the guidance format's an important one. We also want to hear your recommendations on the content in addition to the style and format. If you have a great, plain language way to express something, a good example from any type

of medical product, especially examples of communications and patient input and involvement, we've heard a lot about that today, please submit those.

And it might be a device, drug or biologic product, but we also have a lot of consults and do a lot of work with our friends and food safety and veterinary medicine, so it is open to lots of different products. But please sent it to us. And your experiences and thoughts. That's what's really helpful and moves us forward and it helps shape the guidance that, not only we need right now, but that's going to help it remain relevant as we're moving forward.

So let us know, but I also want to take this as the opportunity to thank the village that got us to this moment. Usually this is when everyone starts packing up, but I want you all to pause as some folks in the last session mentioned to consider how broad and deep the commitment to this effort goes.

Across the FDA, it's not just a couple of pockets here and there, across the entire FDA,

people lent us their best scientists but also their best process people and contractors and a lot of times, we're asked, like, are the clinical review divisions really bought in. And we hope that today you'll have seen that we're at the leadership meetings on a lot of different levels at the primary reviewer meetings that we are all together in patient-focused drug development. It's a key part of our FDA mission.

And while hiring and pay continue to always be concerns, FDA employees are 100 percent mission. So I want to give a few thank yous to Mary Jo Salerno who was just up here who's our project manager who oversees all these different moving parts, the CDER Office of the Center Director, especially, Mina and Meghana, you all have provided so much in the terms of the logistics and policy oversight and getting us and all these documents here today. And that's true especially for Meghana for this entire series that she's worked on.

The Office of Translational Sciences, where I live, our Office of Administrative Operations Travel Team, the Office of New Drugs has the public

meetings team, so you saw them running around all here today and when you signed in, they were there. And a lot of people across a lot of different offices and FDA have supported these guidance efforts. We've had contract supports now for copy editors and technical writings including Rick Turner who's sitting here and helping us edit and shape the workshop document.

CDRH said, here's our digital health team. Here's our patient science and engagement lead. Like, we're going to spend our time doing this and CBER. We have our Science of Patient Engagement leads and other comments and other information from other centers keep coming in because the basis of the document that we had for today's workshop came from the comments FDA's given to sponsors and patient groups.

Over a hundred different reviewers across statisticians and social scientists and data standards experts, clinicians, psychometricians, and more, across all these different therapeutic areas and product types contributed comments that have been written over the last few years and also what they saw

on the horizon.

Several leaders at FDA read all of that, gave a lot of feedback on that. We had a seven-person writing team that took the charge, over a hundred pages of lots of different information and their own expertise to try to put this together with 20-plus people across three centers and a lot of different offices editing and making additional comments to make the document a reality, and we couldn't sustain this effort without the feedback and support of our FDA center directors.

And that also is the feedback and support we've gotten both from patients, industry, academics, and others, many of those organizations that you saw here today. But there are other people that I also want to thank: Michelle Tarver at CDRH, Telba Irony at CBER, and my special thanks to Theresa Mullin at CDER because over 10 years ago, she took this charge and she took the voices of many, internally and externally moved this forward.

And I also want to give my special thanks to Scott Komo. So none of this work would've

happened without his leadership, his staff mentoring, and dedication, the countless hours of public service. And I can fully express the gratitude. Sorry, I told myself I wouldn't cry, but I knew I would. The people who know me, know that. But I can never really express the gratitude he's due because he's been and remains the scientific leader helping the public get that accurate and science-based information they need.

And it's also, I have to thank our families because they put up with a lot of long hours for us to make this happen. But back to what we need from you.

We need you with all the time and care that you can, that means we know what we've done is not perfection and we need help. So we want to hear from you by 11:59 p.m. Eastern. I say that because I was once on the Pacific time zone and didn't get my stuff in on time. So now, I remind people. But we do want, value, and need your input.

So on behalf of the thousands of employees at FDA, thank you for coming today. Send us your comments to the docket. We read them. We want

them written down so we can prepare the best draft guidance possible. And thank you for your continued effort and work in this area. Have a good evening.

(Whereupon, at 6:08 p.m., the proceeding was concluded.)

CERTIFICATE OF NOTARY PUBLIC

I, MICHAEL FARKAS, the officer before whom the foregoing proceedings were taken, do hereby certify that any witness(es) in the foregoing proceedings, prior to testifying, were duly sworn; that the proceedings were recorded by me and thereafter reduced to typewriting by a qualified transcriptionist; that said digital audio recording of 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.

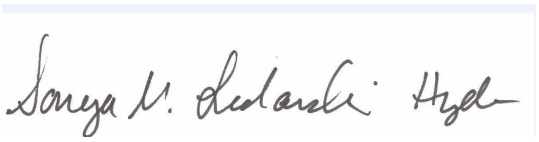
MICHAEL FARKAS

Notary Public in and for the

STATE OF MARYLAND

CERTIFICATE OF TRANSCRIBER

I, SONYA LEDANSKI HYDE, do hereby certify that this transcript was prepared from the digital audio recording of the foregoing proceeding, that said transcript is a true and accurate record of the proceedings 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.



SONYA LEDANSKI HYDE