



Rare Disease Endpoint Advancement Pilot Program Workshop:

Novel Endpoints for Rare Disease Drug Development

A Virtual Public Workshop

June 7-8, 2023 • 1-5 PM ET

Rare Disease Endpoint Advancement Pilot Program Workshop: Novel Endpoints for Rare Disease Drug Development

June 7, 2023, (Day 1) Transcript

*Disclaimer: The following is a transcription of the Rare Disease Endpoint Advancement Pilot Program Workshop. Although the transcription is largely accurate, in some cases it is incomplete or inaccurate due to inaudible passages or transcription errors. It is posted as an aid to understanding the proceedings at the meeting, and the recorded video available [here](#) but statements should not be used for direct attribution without verification and permission from the Duke-Margolis project lead. Please contact Dure Kim, dure.kim@duke.edu, for additional information and assistance.

Mark McClellan:

Hello, I'm Mark McClellan. I'm the director of the Duke-Margolis Center for Health Policy. On behalf of the center, I'd like to welcome you to today's workshop and tomorrow's workshop on Rare Disease Endpoint Advancement Pilot Program for Novel Endpoints for Rare Disease Drug Development, which is being convened by Duke-Margolis and our partners at the Food and Drug Administration. This workshop will illustrate challenges and opportunities in rare disease endpoint development, introduce stakeholders to the RDEA pilot program, and highlight how RDEA is structured to support sponsors who may encounter challenges with endpoint development.

At this meeting you'll hear from a variety of speakers about rare disease endpoint development examples to gain a better understanding of endpoint development challenges and opportunities that may be cross-cutting or distinct in different areas. Workshop programming will also facilitate, we hope, a shared understanding of the RDEA pilot program's purpose and structure, including its key features such as sponsor disclosure requirements. Learnings from other FDA pilot programs that share common features with the new RDEA pilot program are also going to be part of this discussion. So the event is intended to serve as a resource for sponsors and other attendees interested in learning how they might engage with the FDA through this new program, building on all this other related experience.

Before we begin, I want to go over a few logistical and background considerations for our discussion as well as some housekeeping notes. So on the next slide, this is a short note about our commitment at Duke University to academic independence and a diversity of perspectives. We do not take partisan positions in our work, we do encourage frank discussion from a wide range of views to help inform constructive discussion of the topic and you can find out more about our policy at links on this slide. Also want to disclose that I am an independent director on the Board of Alignment Healthcare, Cigna and Johnson and Johnson for my own background. And again, we've got a wide range of other perspectives here as well.

Onto the next slide, here are a few reminders about our meeting format that will help us organize the discussion in this virtual setting. Just a reminder that all the meeting materials are available on the Duke-Margolis event website. And for our registered stakeholders, if you would like to ask a question during the presentations and open discussion portion of the meeting, you can enter it at any time into the

Zoom Q and A box. That should be at the bottom of your screen. Please be aware that we do have a significant number of meeting participants that we may not be able to get to all the questions, but we may also have some opportunities to certainly appreciate the comments that come in and also maybe answer some of those I am writing out live as well.

A recording of this meeting is going to be available on the Duke-Margolis event website and our YouTube channel after the conclusion of the event. Finally, we want everyone to know that Duke-Margolis is convening today's meeting in cooperation with the FDA, but this is not a Federal Advisory Committee meeting. We are very interested in your viewpoints and input on the topics discussed today, but we won't be following Advisory Committee procedures to try to get to a consensus on results. We won't be taking any votes. The meeting instead will be a success if there is a frank and constructive exchange of views and ideas and open discussion on the topics for the meeting to help advance how we're able to work together on these topics. We're thrilled that you're able to join us today and I want to thank you in advance for what I expect to be a timely, interesting and productive discussion.

So let me turn to what exactly is involved in that set of discussions. So I'd like to go through the agenda. We're meeting over two days, as you know. For day one of our workshop, we're going to start with opening remarks from our colleagues at FDA, followed by four sessions focused on considerations in developing rare disease endpoints. Session one will cover digital health technology, session two will cover biomarker surrogate endpoints. Session three will cover clinical outcome assessment, and session four will cover multiple endpoints with a focus on multi-component endpoints. And we're aiming to adjourn today's meeting before 5:00 PM Eastern time.

Then we'll be back for day two. So I'd like to go through that agenda as well. For day two, our speakers will introduce attendees to the RDEA pilot program and provide some additional information about the process and required elements for submitting a proposal. Session five will provide an overview of the RDEA pilot program. Session six will provide an overview of the RDEA pilot program processes. Session seven will cover elements of RDEA proposals and meetings, session eight will be a moderated Q and A and discussion of these topics.

And then in session nine we'll turn to covering experiences and lessons learned from other PDUFA pilot meeting programs that could be relevant for making this program as successful and effective as possible. Session 10 is reserved for public comments to make sure we're not missing out on any other ideas and perspectives that participants want to bring and attendees want to bring to the meeting. And then we're going to be aiming to adjourn day two at 4:45 PM Eastern time on Thursday.

Nancy Allen Lapointe:

Great. Thank you so much, Mark, for that introduction to the workshop and for providing an overview of the agenda for today and tomorrow. My name is Nancy Allen Lapointe. I'm a faculty fellow at the Duke-Margolis Center for Health Policy and I have the pleasure of introducing Doctors Peter Stein and Celia Witten for opening remarks from the FDA. Dr. Stein is the director of the Office of New Drugs in the FDA Center for Drug Evaluation Research and Dr. Witten is the deputy director of the FDA Center for Biologics Evaluation and Research. So first off, I'd like to invite Dr. Stein to provide his opening remarks.

Peter Stein:

Nancy, thanks very much, and Mark as well, for that introduction to this, what looks to be a terrific workshop. On behalf of the FDA Center for Drug Evaluation and Research, I'd like to welcome you to the Rare Disease Endpoint Advancement Pilot Program Workshop on Novel Endpoints for Rare Disease Drug Development. CDER has a long-standing commitment to the development of therapies for rare diseases. This commitment is exemplified by this PDUFA VII Rare Disease Endpoint Advancement pilot program

and associated initiatives. The Rare Disease Endpoint Advancement pilot program, we use the term RDEA for short, R-D-E-A. FDA has had success with previous pilot programs such as the PDUFA VI pilot programs for model-informed drug development and complex innovative trial design and looks forward to what this program can do for rare disease drug development in the challenging area of endpoint design and selection.

The lack of regulatory precedent, small trial populations, and often a limited understanding of disease natural history for rare diseases creates unique challenges in identifying appropriate efficacy endpoints for clinical trials intended to evaluate the effectiveness of therapies for rare diseases. And often small sample sizes with limited study power requires sensitive endpoints with as little variability as possible, another challenge. While FDA has several initiatives that could potentially benefit rare disease drug development, the RDEA program will focus specifically on supporting this specific challenging area of rare disease drug development, developing clinically meaningful and measurable endpoints for clinical trials of rare disease therapies.

Developing rare disease endpoints is truly an interdisciplinary activity, requiring expertise in our Division of Rare Disease and Medical Genetics, in our statistical group, in our clinical outcome assessment group, and in divisions across OND and offices across CDER and CBER. FDA will deploy the expertise of our interdisciplinary experts as needed to collaborate in developing rare disease trial endpoints with sponsors admitted into this RDEA pilot program. The novel endpoints developed through RDEA may be presented by FDA such as in guidance documents or on a public facing website or public workshops as case studies, including prior to FDA's approval for the drug studied in the trial. Before FDA grants the initial RDEA meeting, the agency and the sponsor will agree on the information that FDA may share publicly in these case studies. When feasible, FDA will notify a sponsor in advance when the sponsor's program is the planned focus of a public discussion. FDA looks forward to sharing learnings from the RDEA pilot program with rare disease stakeholders in many ways, including CDER's Accelerating Rare Disease Cures program, the ARC program. But most importantly FDA looks forward to the program's benefits in moving closer towards treatments for rare disease patients who currently have no available treatment options.

Well, thank you for your active participation in this meeting and I'm looking forward to the many learnings that will come from this two-day session. Thank you very much.

Nancy Allen Lapointe:

Great. Thank you so much, Dr. Stein, for those opening remarks. Next, I'd like to invite Dr. Witten to provide her opening remarks.

Celia Witten:

Thank you. I'd like to add my welcome to this workshop on behalf of the FDA Center for Biologics Evaluation and Research. There has been an explosion of interest in a number of notable successes in development of therapies for rare diseases in recent years, and CBER looks forward to seeing those successes continue. In particular, there have been some particularly remarkable developments in the area of gene therapies for rare diseases. The development of appropriate endpoints is critical for success. There are limited precedents for some diseases and later in this workshop you will hear about the approach one sponsor took to develop a clinical outcome assessment in the case of a gene therapy for an inherited retinal dystrophy.

I want to take this opportunity just to mention that we recognize the importance of patient engagement early in drug development, including the area of clinical endpoints development. The RDEA pilot program will provide additional support to sponsors in developing novel endpoints for the product

development and we will look forward to sharing FDA experiences and knowledge gained from this program. And thank you very much. And now I'll turn it back over to you, Nancy.

Nancy Allen Lapointe:

Great, thank you so much Dr. Witten. Now we're going to move into our first session. So I'd like to invite the moderator of that session, Michelle Campbell, to take it from here.

Michelle Campbell:

Well, thank you Nancy and thank you Dr. Stein and Dr. Witten for this great opening remarks and really in setting the stage of a very dynamic two-day meeting that we're about to have to explain the RDEA project. And as a reminder, stakeholders may provide written comments regarding this, as that slide just said. For those of you who may not know me, my name is Michelle Campbell and I'm the associate director for Stakeholder Engagement and Clinical Outcomes in the Office of Neuroscience in the Office of New Drugs.

I would like to welcome you to our first session of the day entitled Considerations in Developing Rare Disease Endpoints: Digital Health Technologies. In this session, we will focus on exploring an example of a digital health technology, and attendees will learn about the relevant endpoint development challenges and opportunities when using DHTs. The session will end with a panel discussion where the panelists will share their thoughts and opportunities and challenges when developing rare disease endpoints that involve these DHTs. I do think we will have a very dynamic discussion today. But before I introduce our next presenter, I want to share some relevant background information. Next slide, please.

Recently, Duke-Margolis also hosted a meeting regarding digital health technologies in March of 2023. And just to set the stage on what a digital health technology is, it is defined as a system that uses computer platforms, connectivity, software and sensors for healthcare and other related uses. And as you can see on the slide, it covers a broad range of potential uses for clinical trials and medical product development. Next slide.

Additionally, there's a large range of what these can be and so as these are just some examples of what a DHT could potentially be used for, they range from something as simple as electronic data collection of a patient reported outcome or other clinical outcome assessments, to continuous glucose monitoring, to actigraphy and accelerometers, to even portable EEGs. So as you can see, we do have a vast wealth of information and knowledge and devices for DHTs that are coming out from the software that could help collect additional data and add to the patient experience of their lived experience of their disease and disorder. Next slide, please.

It is also important to remember that DHTs need to be fit-for-purpose for use in clinical investigations. And if folks are familiar with the development of clinical outcome assessments, we often talk about fit-for-purpose there. But digital health technologies should have those same considerations, and what we mean by this is that we have a characteristic event or a clinical event we want to study, that the DHT can measure what is of interest, that we have the right population, we have the right technology. And that is not just the technology that's being used to capture the event, but also that the person using it, such as patients, have an understanding of what they're supposed to be doing and if there's any training that needs to be done for them to understand the technology that they need to capture the data on, as well as is the DHT design and does it operate in a functional way to collect the data? So next slide, please.

Next slide. So I would like to turn this over next to our presentation. For this session, I'd like to introduce Dr. Laurent Servais who is a professor of pediatric neuromuscular diseases at the MDUK Oxford Neuromuscular Center and an invited professor of child neurology at Liège University and his

presentation on work on developing a digital health technology for Duchenne and muscular dystrophy. Lauren.

Laurent Servais:

Thank you very much Michelle. I'm really appreciating the kind introduction. May I have the next slide please? So I think that I don't need to inform the audience that we have a huge attrition rate in clinical development and the cost of clinical development has been exploding with years. So the key question for me as a physician running clinical trials is how could we make clinical trials more efficient and how could we ensure that we get the right answer from the trials? And of course we can make trials longer, we can use more patients, but this is not possible in rare disease. So probably the only way is to have better outcomes that are more sensible, more sensitive to change. May I have next slide?

And this is from a study done nine years ago during which we were recording six-minute walking test and at the same time the patient was wearing an activity monitor. And we could figure out that when the patient was so happy to leave the hospital, he was walking much faster than during the six-minute walking test, which means that we were believing that we were recording the maximum velocity of the patient during six-minute walking test, but we were not actually. Next slide.

Yes, you can continue to next slide. And another feature is what I call the Eiffel Tower paradox. Next slide. Even if the patient does his best to do exactly what we ask, the very first day I was a baby biologist in Paris, I saw a patient was performing very poorly in the six-minute walking test and I was so sad for him. And mom told me, "Oh, don't worry, it's just that yesterday we went to visit Eiffel Tower." Because he had rare disease, he was living far away of Paris countryside. So the first thing he did when he came to Paris was to visit Eiffel Tower and then Euro Disney. And of course he was exhausted. So we need to figure out that in the real life we're gambling decades of research and billions of investment on the outcomes from patients who have just visited the Eiffel Tower or Euro Disney. Next slide.

There are more and more digital technology and wearable devices that are used by years and this is the recent papers that shows for several neurological disease, the expansion with years. Next slide. Especially for motor tracking, most of affordable technologies is used for motor tracking, a little bit less certainly for sleep tracking, cognition tracking, or speech tracking. Next slide.

This is from a paper that is currently in minor revision. We have reviewed all the studies that have included in rare disease wearable technology and we found 56 studies that have included 3,605 patients. Next slide. And as you can see, the different wearable sensors were placed at different part of the bodies and patients were wearing between one and six or seven sensors. So obviously, yes, there are many studies using wearable technologies, but at the end of the day, today there is no wearable outcome or digital outcome that are qualified by FDA or by EMA as a primary endpoint. Next slide.

And I think that the explanations resides in what I call the Christmas gift paradox. What is the Christmas gift paradox? Next slide. As a researcher I received for my Christmas as a Christmas gift from my wife, from my mom, I receive a Fitbit, an Apple Watch or a beautiful tube. And immediately, as in my clinical trials, I will think, "Well, what can I do with that?" Next slide. And actually the short answer is, "Nothing." Next slide. Because the problem is that these devices are not done, are fit-for-purpose to conduct clinical trials. They are very nice tools that you can offer for Christmas. They are not tools that you can really use in the context of a clinical trial, next slide, especially the fact that they are not calibrated for temperature and then when the patient goes out and in and there is huge change of temperature, then you can quickly be into the wild.

So I did the same mistake 12 years ago trying to start with a very beautiful and very nice device out of the shelf, and trying to get something that makes sense for my patients. And unfortunately it did not work. Next slide. Then we started having a discussion with the engineers and the doctor asked to

engineer, may I have next slide? And the animation... Next slide, please, move ahead, yeah. I would like an altar of movement. That was my dream, to have an altar of movement. And of course the engineer did not exactly understand and then I start explaining what I wanted. Next slide. I want you to be able to identify all the movements of the patients, next slide, and to be able to identify them very precisely. Next slide. And I want to do it in an uncontrolled environment. Next slide. And I want to do it during two years without any shift of the time period. Next slide. And I want to be compliant of course with data security and all regulations around the world.

And as you might imagine, 12 years ago the engineer answered, next slide, "Are you sure that the moon is not enough?" Because this was indeed quite challenging. Next slide. So you can please move in the animation. So the question was much more what do we need as a clinical trialist? So we started not with a beautiful device but with the prototypes, it was terrible, it was not wireless at all, but we could identify the right sensors and how to calibrate it. Next slide. It took some times, but then we went to a second prototype that was very big and bulky and it was Bluetooth with big batteries. Next slide. And then we could move one year later with something that was still ugly but a little bit less ugly and we could start recording patients. Next slide.

It took years until we got the medical device that was still ugly. Next slide. And then we started to get interesting outcome as I will describe for Duchenne, next slide. And we could finally move to a device that was much more acceptable and beautiful for our patients, 11 years after having started, next slide, and get something that could be potentially qualified as an outcome at the European Medicines Agency level. So just consider 11 years to really get a medical device that is fit-for-purpose in which you understand every single component and data transmission and on which you can really rely in the context of the clinical trials. Next slide.

It was very long. And can you play the little video? This is the type of reconstruction that we can do from the stride. Look, the patient falls. So not only you can measure the strides and identify every single stride and measure every single stride, but you can also identify the falls that the patients stand up and walked again. So as you might imagine as a neurologist, when you have such a tool, you just want to play with it. Next slide, please.

So you have the device, it is compliant, it is beautiful, the patients likes it. That's fine. So the patients wear the device during two weeks and what you get a mountain of data and that's where the play begins. What can we do with this mountain of data? But actually it depends off what are the symptoms that you want to identify and what are the objective. The symptoms could be, "I want to identify the power of how fast my patient climbs stairs or the fatigue or the activity or the falls or the ataxia or the abnormal movements." And then the question comes, for which purpose? Next slide. Do I need for instance to be sure that my drug works on limited number of patients even if it's not clinically significant to take my go/no-go decision?

Next slide. Do I need to convince FDA during the pivotal trials? Next slide. So do I need to convince the payers that it change the life of the patients? Next slide. Or do I need to convince my patients that it should continue to take a medication and to move to a individualized medicine? So that's the questions that we need to ask ourself, "How do I extract from this mountain of data," next slide, "a single variable? The holy grail?" Next slide. And a single variable that is validated, well-understood, and that I can enter into a database. And this holy grail for Duchenne muscular dystrophies, probably the stride velocity 95th centile may have, next slide, that is on the edge of being qualified as a primary endpoint in Duchenne muscular dystrophy by the European Medical Agency. May I have next slide? Yes, you can move next slide. Thank you so much.

So what is it? How does it work? So the patients wear the device on the ankle, both ankles or ankles halfway or width are raised as you want, as we want, actually. Next slide. During the night the data goes

into the cloud. Next slide. Then we identify every single stride of the patients. Next slide. We measure the strides and we can make distribution of the stride velocity or the stride length, and we can select the 95th centile, which is the top performance of the patients at home. Next slide. Okay, please move to next slide.

So it represents actually the old measure top performance of the patients. And we know that first it's important for the patients and that our patients with Duchenne muscular dystrophy, they complain about not being able to run with the others, to follow the others when they are doing some work. And we also know as physicians that the very first thing that our patients lose is their top performance. Next slide. Next slide. So then why the 95th centile? Actually we computed from several natural history studies what is the percentile that correspond with the best power to measure a patient's change. And as you can see, if you go to the top performance, you get a better SCRM. The problem if you go to a 99 centile is that you have a very limited number of strides and high variability. So there is a kind of sweet spot and this sweet spot is about the 95th centile.

Next slide please. So we could estimate that, next slide, rather than using during one year 112 patients, we could use a much smaller number of patients during six months, next slide, which is about 14. And that was in the secondary qualifications at the EMA in 2019. Next slide. So of course you can see while the problem is the variability of actimetry. Next slide. And I've got two theories. When I was working in Paris I used to say when the French are on strike, they walk more because there is no public transportation. Next slide. And now I work in the UK and I used to say when it pours, British people walk less, and when it simply rains, they walk more. So there is a huge variability of actimetry. And this is true, but next slide, if you are interested not only in number of strides but all the patients actually walk, so the stride velocity, it is much less dependent of the weather, for instance.

And then another key feature is that it really depends how long you measure your patients. If this is for instance, an activity parameters of a child, and you can move with the animation please, you'll see that there is a huge variability. If the patient live with the mother next slide or during the weekend, and then the next week when he is with the father, next slides, please move the animations, we will see that indeed there is a huge variability. But if we average the signal on a three-day period, next slide, you will decrease the variability. And if you continue to average on longer period, next slide, please, you will actually decrease the variability. So the variability depends actually on the duration during which you average a signal. Next slide.

And this is what we call in mathematic the Allan variance. And that helped to define the sweet spot for Duchenne that was between 50 and 180 hours to get the best variability, which was only 3%, which is quite low. Next slide. So we had then to understand what is the influence of compliance on the stride velocity 95th centile, and there is no influence, good news. Next slide. We had to have of course a control match population and to show that there is a good discriminance between patients and controls, next slide, that the patients in which we start steroids, in which we start treatment, are actually improving. Some of the patients who are steroids for a long time are declining.

Next slide. We had of course to correlate this new outcome with the existing standard. So for the clinical consistency, next slide. And of course to define the minimally clinically important difference, which is not easy. And as you know there are several meters, but it's beyond the scope of this talk. Next slide. So with all this data in 2019, we could gather the secondary qualification in EMA and we gather many more data and many more analysis. And the CHMP of EMA issued the drug qualification opinion as a primary endpoint. And as you can see the end, the public consultation is now over and we have answered all questions. So we are quite confident that this could be finally the first digital outcome ever approved by a regulatory agency. Next slide.

So this has been a shift with the collaboration of a broad community, next slide, including patients advocacy group and several important companies. Next slide. Nowadays, this type of technology is used by 25 sponsors in 10 different conditions. Next slide. And we have over 1 million to a hundred thousands hours of recording. So the takeoff message is one, please move ahead, digital outcome has made its way in the regulatory landscape and it has the potential to reduce the duration and the size of clinical trial. We're now using it in Duchenne, in SMA, FSHD, but also in non-neuromuscular condition line, multiple sclerosis, Angelman, Dup 15q. But again, we need to find the right outcome for every single condition to be sure that the outcome that we use really represent and means something for the patients population. Next slide.

So the key learning from all these years for me is the 3D rule. Next slide. Rule number one, the quality of the device is key. Rule number two, we need to develop interaction between engineers, MD, the community and the regulators. Next slide. And the third D is the early and high quality of data collection. We need data and this is not easy, certainly. Next slide.

And our key question, and probably we're going to discuss some of them, next slide. The first one is how to make an outcome measure development really attractive for industry and investors, because definitely the return is not the same as for a drug. Next slide. How to deal with less common disease. Can we extend the qualification from one disease to another context of use, like for instance to limb-girdle muscular dystrophy? This is a completely unbeaten track to my best knowledge. Next slide. And the third important question is how can we assess the difference in the process between FDA and EMA? Because this makes the qualification very time- and energy consuming because we have to follow two different processes. Next slide.

Laurent Servais:

We have to follow two different processes. Next slide. Finally, how can a qualified outcome evaluate with time? And when something is better, how can we do? Next slide.

Finally, I just want to thank the team, the team in Liege and my PhD student Margaux. Next slide.

And a great team of Sysnav engineers, especially a very specific word, next slide, for Melanie Annoussamy, who made a fantastic work and unfortunately passed away two weeks ago. She was a brilliant scientist and I dedicate this talk to her. Thank you so much for your attention, and happy to participate now to the discussion.

Michelle Campbell:

Thank you, Laurent, for that discussion. I do invite my panelists to turn on their cameras and to begin our discussion. I also encourage the audience to put some questions in the Q&A if you have questions, as we have some time to have a very thoughtful discussion. But as we start, I'd like to introduce our panelists with us.

So we have first Dr. Damien Eggenspieler, who serves as the Head of Healthcare Activities for Sysnav. We have Hussein Ezzeldin, who is the Acting TL in our Center for Biologics, leading Digital Health Review Team there, and all the digital health technologies that CBER encounters. I'd like to introduce my colleague Dr. Ami Mankodi, who is a reviewer in the Division of Neurology 1 for neuromuscular diseases and disorders. I'd like to introduce Leonard Sacks, who is the Associate Director of Clinical Methodology in the Office of Medical Policy in CDER. And then Laurent will be joining us for this panel. So thank you, everyone.

So, I really appreciate the presentation, and the presentation that you said, and really the lessons learned from this. As an innovator and earlier adopter of using DHT, I think is always great to hear from

those who kind of started it, the process, the lessons learned, the growing pains, as you were exploring the developments of the use of the actigraphy, and what was 95 stride velocity, and that process of getting evaluated by a health regulatory agency. Next slide, please.

So I was hoping that we could talk about, start with is, really, what are our biggest challenges that stakeholders experience when developing DHTs for rare disease drug development? And what are maybe some from those lessons learned effective strategies that you would suggest that others consider when taking this approach in the development, and really thinking about how can this support a clinical trial endpoint, and be informative to an overall, potentially a marketing application that would come in for review?

So Laurent, you just gave a great presentation on a lot of the challenges, but is there anything else you would want to add from your experience, or advice you would give folks who want to start doing this that you've learned from going through a review process through EMA?

Laurent Servais:

Yeah. I think the main challenge is first to get data, because when you come with your terrible ugly device, and you want to go in a trial, the sponsor will say, "Wow, I won't ask the patients to wear it." But on the other hand, it's very difficult to fund a very beautiful and acceptable device if you don't have data at first. So, getting data is a big challenge, and having a development in which you can in parallel to progress on the hardware, and the software, and acquiring data, is really, really challenging.

The second thing is that our experience with the EMA, for instance, when we came in 2016 was difficult because obviously this was not a very clever beaten track, and it was quite new, and it was difficult to find the exact pathway. Our experience for the primary endpoint qualification was much more straightforward, because since then I think that everyone has understood the limitations, the strength, the byways, the pitfalls of digital technology, and it made our life much more easy.

So I think that for me, again, the main challenge is access to data. How can we quickly get data using this type of technology? And the second main challenge was certainly the interaction with the regulators, that was challenging, let's say, at the very first. And now we are facing new challenges. And the new challenges is, how can I do the same for limb-girdle muscular dystrophy type 2Z, in which there are 15 patients in the US? How can I extrapolate to new population?

The second challenge is, how can I make my outcome work better, and getting the same level of qualification? How can make my outcome better and keep my qualification? So all of this, which is a moving target, how can it evaluate and keep the qualification? These are, at least to me, still open questions.

Michelle Campbell:

Well, thank you for that. And that is I think some really great information, and I know I wanted to follow up with you on a couple of questions about the data, in terms so you can start thinking about maybe a response. But I do want to get into asking a question a lot of folks are asking, and a question that we're also faced with that challenge is, how do we take all of that raw data to determine what is that cut point to decide that this is what's going to be supporting the endpoint? So, I want to get to back to that conversation in a second, but I want to ask Damien, as part of that device or technology group that's working on that, what would you see that some of the challenges that you have seen so far as the development was going on with the Actimyo?

Damien Eggenspieler:

Thanks a lot. And again, the answer is not going to be vastly different from what Laurent said. I think one of the things that is important is long-term commitments, because when you start to go into this adventure, you need to understand that you're trying to climb the north face of a mountain. And there's a lot of left and right things that might be attractive business points, but that for patients in the end will not make sense. So I think this long-term commitment trying to find the right partners in patient associations, because in the end, if you're grounded by patient associations, they're never going to stop to bug you until they have something that works for the disease. So I think we've been seeing a lot through the driving force from patient associations.

Collecting data is super important, obviously, because we all have a lot of cognition bias, because we all use our watches every day, and there's a lot of people who would tell you that we would need to go into this direction, this direction and so on. But in the end, our decision needs to be grounded on data. And at the beginning it's very difficult to acquire data, to get access to the data, especially because you're piggybacking a lot of those data on drug development.

And indeed, the last bit is, how do you validate, not necessarily the endpoint, but first of all the analytical validity of your digital device? So, for us on the ambulatory patients, it was really important to have meaningful anatomic variables, and this anatomic variable is really the steps. And we needed to measure the steps with high precision, because we know that once we do that, we know the answer is in the dataset that you have, the mountain of data that Laurent was describing. And then it's just a way of looking at it, a way of trying to dig into this data, trying to understand that. Numbers of steps might not be important because you have a lot of environmental issues, or bias that might bias your data and so on. So then it's just a matter of repeatedly trying to dig into this data, both with the discussion with patients, with discussion with your data, and as well trying to really understand the deep down of the meaningfulness of those data.

Michelle Campbell:

Yeah, thank you for that. And I'm really glad you brought up the analytical validation aspect of digital health technologies, because it's not just capturing and thinking about what do we want to measure, and what is this concept, and is this meaningful to patients, but it's also, does this technology work? Can we replicate it? Can we understand it? Can we interpret it? And I think that is an important thing. And obviously, FDA, we do have our guidance out on digital health technologies that does talk about this information, and I think that is a really critical thing that has to be thought about if one is the plan to use a digital health technology in their clinical trial.

I want to turn it over to my FDA colleagues with the same question. And with the mindset of us, as a regulatory agency, some of us are primary viewers, what are some of the things that we're seeing, challenges that we're seeing, and what advice would you think about that we could provide to help people think about this early, perhaps in their development program and before they come in and get some advice, what are some of the challenges that we're seeing that maybe we need to think about a little bit more in a way that can be beneficial in the long run to help support our clinical trial endpoint hierarchies?

So Leonard, I was wondering what your thoughts are coming really more from the Office of Medical Policy, and really having to help really lead in the efforts of DHT in CDER. What are some of your thoughts with that that we've been seeing, or that you see, are some of the challenges right now?

Leonard Sacks:

Thanks very much, Michelle. I think the most striking challenge to me is just time to develop confidence in new measurements. And I think that's something that affects all of us. It's the entire ecosystem, it's

regulators, it's drug developers, it's physicians and it's patients. And I think the only way that we are going to get there is by developing a good body of data and experience. And I think that's being done in a very piecemeal way at the moment. I don't want to tag onto another challenge which Laurent mentioned, and I really thought it was a very good observation, is partnering with engineers because I think there's obviously a temptation to look at devices that are around, and easy to use, and available commercially like smart watches and Fitbits. But really we are in an area of sophisticated medical science. It seems a shame that there isn't a specialty of device engineering for measurements.

And I think, as you look at different diseases, there are different parameters and different features that you want to measure. And I think that seems to me to be an additional challenge. I mean, we know that consumer devices provide some information. I think there's a fair amount of skepticism in many circles that that's not adequate. And so I think the quality of the measurements is one thing that's a challenge, and a part and parcel with that is developing the right sources of data. And I've been sort of incubating the idea that really any clinical trial in diseases which are suitable for these types of measurements should become a kind of piggyback trial where measurements should be made in parallel using digital health technologies in traditional ways, for us to be able to benchmark how these technologies look against other technologies. And I think those are some of the principle challenges that occurred to me. Thanks.

Michelle Campbell:

Thank you, Leonard. And I know that we every day are discussing these challenges and the opportunities, and what is the science that we need, and how can we work collectively to help advance the science. So Ami, I'm going to turn to you. You are a medical officer, you are reviewing every day applications, INDs for advice. And as you're seeing things that are coming in, and we do know that in the Office of Neuroscience we do see a dynamic utility of the use of digital health technologies. What are some of the challenges that you're seeing from your side as that reviewer with DHTs coming in, and where do we see that maybe we need to go and learn from those challenges that can help us, again, for some of our clinical trial endpoints?

Ami Mankodi:

Thank you, Michelle. And it's been really exciting opportunity to be part of this panel and also understand from Laurent, actually, what the process was, because he touched upon lot of challenges and how he overcame in his journey. So I think we are all excited about the technology revolution. We are also palpating the enthusiasm that technology big digital or otherwise, I think it's geared towards better capture of patient's voice in drug development. Patient-centric approach with focus on health equity is what we promote in our division and across the divisions. So engaging patients earlier in the process. So here for clarity of communications, clinical outcome assessment is a measure that describes or reflects how an individual feels, functions, or survives. And there are different ways of measuring that. Most of us know patient reported outcome measure. The one that we learned today about the ankle stride velocity would be clinical outcome measure performance based, but there are patient active voice. So patient reported outcome measure, there's clinician reported outcome measure, observer. So caregiver reported outcome measure. The proxy reporting is something that we discourage, and I really want to highlight that endpoint is something that is designed, it's based on clinical outcome assessment, or otherwise be digitally measured. Digital is just an approach, and that is intended effect of the drug measurement. So be it efficacy, benefit or risk during the trial. So it's very, very important to define whether it is clinically meaningful reflection of how patient feels, functions, and survives. So for digital technology, or otherwise, it should directly measure what is a part of individual's daily life. It should be important and meaningful to patients in their daily life. It should be sensitive or responsive to a known

change in the concept of interest, be it response to the drug for a trial, should occur at that frequency because the trial duration is limited.

And I should also emphasize that the roadmap for these outcomes are still the same. So first is defining what is relevant to the disease, and two, the patients, so concept of interest, understanding the disease. So natural histories are not known for many of the rare diseases. There are small populations, so sometimes very, very small populations that are scattered in remote areas, geographical barriers, but there is also digital barriers that will have to overcome. So access is one thing, literacy is one thing, patient's choice also will matter. So we need to engage the patient populations. Thankfully, in rare diseases overall, there is more, I should say participation by patients, and community in general with the researchers, with the clinicians, with the patient foundations, and with regulators through the listening that we have for the patient listening meetings that we go to. I would also say that it's challenging to specify the fit for purpose, Michelle, that you mentioned, for the clinical trial, from the clinical outcome assessment.

So, how do you conceptualize clinical benefit for the trial? Whether it'll be used in entry criteria for stratification, for the disease population in the study. What is the definition for the endpoint, the position of the endpoint? Primary, secondary, et cetera. And then determining whether the context of the use, so validity of the tool, content validity, construct validity, face validity for the use in the clinical trial. Is it sensitive, responsive to change, and the change is meaningful to the patients? And I think we encourage participation right from the pre-IND state. So because I'm from the review division, but there are other means like drug development tools, for example, there's a critical path that we partner with. So in the pre-IND, right from the pre-IND, and then continue through the life cycle of the drug development as a review team, we would like to know more about the rare disease, what are the knowledge gaps, what are the steps taken to address those knowledge gaps.

So how do we address the partner in this journey? What are the proposed trial designs, the endpoints, consideration of novel endpoints? Why those endpoints are fit for purpose for that given application? What is the intended study population? Can we broaden it? In future, what are the plans to broaden it to a different population? Whether the results can be generalized, statistical analysis plans. So one thing that we look at is approach is taken to address missing data, different types of biases that is involved. So there are a lot of different aspects, and I should also emphasize that we work together as a consulting, I'm a clinician, so I'll just take the example of a hospital where we consult different teams. So we work with different divisions across the agency. And it's like a piece of puzzle, so expertise come from all different sections in order to make the project patient's voice for patients, and ultimately make an impact in their quality of life.

Michelle Campbell:

Thank you, Ami. I'm going to turn it over to Hussein, and I really want to thank you for being here really for our Center for Biologics, because this is a joint CDER-CBER pilot project. So I wanted to get your thoughts really from the CBER side, particularly since your therapies that may be coming in are for cell and gene-based therapies. What are some of the things that need to be considered with using a DHT in the CBER space?

Hussein Ezzeldin:

Thank you so much, Michelle. I think my colleagues really touched upon a lot of things, but maybe what I would like to just highlight from Laurent's presentation is that the amount of time and effort it took the development team and the clinical team to be able to arrive to a device that they can trust, and they can have confidence in the measurements, and identify that the correct parameters and outcomes that they

are able to point to and use as an endpoint is really important. And I think Leonard touched upon this, is that over the shelf devices, as much as they might be promising, they are really not very adequate for the clinical use. And I myself lead a natural history study in a rare disease, and I can tell you that it's not easy to have the patients and their caregivers that are already dealing with the huge burden of the disease, and trying to get by on their daily basis to add more things for them to use.

Especially, when you are speaking about exploratory endpoints, which is really important, that this is something that, as Ami mentioned, we want to be really involved from the beginning even on very early stages for exploratory endpoints. We want to be there, look at how the sponsor is thinking about capturing this data, how can we look at it from a statistical point of view. But my experience says that you really want to think about how to remove the friction. So, how you can make this technology easy for the patients to use, and you want to make sure that they have the appropriate training on how to use these devices. We all know how the technologies can be very finicky with very small changes from the subject or the patient side. So you want to have the correct training on how they can use the device, so you have high confidence in the data that you're collecting.

Also, when you look at rare diseases, you want to be able to monitor the patients for a specific period of time. And you want to make sure that this device is functioning the way you are expect to function. Laurent showed data that was looking at maybe two weeks, six months, things like this. And this is the kind of data that we want to be able to see, especially in the development phases, to have this confidence in the data. This is really important. And of course, having access to the raw data prior to doing some analysis is really important to be able to identify any issues with the technologies, because we all know whether it's a consumer base or not, the technologies sometimes have some hiccups in them. So I think it's really important that we have access to the raw data, and look at it and be able to analyze this as well.

Michelle Campbell:

Thank you for that. And I like how you mentioned about the hiccups in technology, because as was shown in earlier slide, the variations of what this can be of that across technologies. It could be even something for collecting electronic data of the PRO, for example, and that technology that's being used on tablet, as an example of there could be a hiccup and a backup to what happens if we did not know the correct charging of maybe the device, what is that charging time to get full capacity to capture data. Or if there was a software update, I think those are some examples that we often hear about. So I really appreciate you bringing up that point. I want to thank everyone for putting in some really great questions in the Q&A. And we have about 15 minutes left, and I know we're not going to be able to get to all of them, but I'm going to let my panelists at least know ahead of time my order I'm about to come to talk to you.

So I'm going to start with Damien with a question, and then I'm going to go to Laurent, and then I have a question for Leonard. So you guys are, they're the order I'm going to come with. But Damien, we've been getting a lot of questions in talking about Sysnav in the data collection of the work to pull that data. There was a lot of great discussion about getting data was hard, and so did Sysnav have to sponsor your own trials, or were you able to work with industry to be able to put a collect data through maybe an ongoing clinical trial, or any type of natural history studies? And then, what type of partnerships or collaborations occurred over the course of this development with the various stakeholders in drug development to help also in that data collection aspect?

Damien Eggenpieler:

Thanks a lot for the question. I think the reality is that we need both. So we need both to have our data, and to generate our own data, for several reasons. For data access, but as well for the scientific rationale of it, because some of the protocol design features might be specific to what you're trying to show. So for instance, how much time do we need to average the signal on the slide that Laurent showed. We need to have recording periods with the device over a year. So to really cover the full fledged of a year, and then going with the sliding windows and trying to understand what were the different liabilities of these measures and so on. So there are again scientific questions that can only be answered by specific protocol. Then, at the beginning it's quite easy to prime your data collections with small size data, and that oftentimes goes through patient associations through the different type of routes.

We need as well to do a lot of the control environment to prove the analytical validity. And I think that those type of data are best hold if Sysnav is actually sponsoring those data. So that would be a set of data that I think it would be natural for us to collect. And then, since we are talking about rare disease today and tomorrow, for some of those disease, there's not enough patients that match the criteria for us to do large scale data. So we need to rely on drug developers, we need to rely on... and again, the question is, how much data do you need on your own to get to people to trust you? And that's again, I guess what you mentioned, Leonard, it's getting to this point where people are confident that they're going to record data that are meaningful. And when I say people, it's A, drug developers, as well patients.

So to get patients involved, we know at the beginning we have an ugly device, but we were able to show to the patients by showing them what we are recording, by showing them what we could see, and so on. We're able to get their trust. And they prefer to wear a device that is poorly designed in terms of the shape and so on, but that is meaningful, rather than to go for the pain of going every day to wear a device where they don't know what the outcome will be. So I think you need to build trust on the different stakeholders to really get everyone on board this adventure. And this involves indeed getting data from drug developers.

Michelle Campbell:

Thank you for that. And I think that just highlights why we have to work together when it comes to digital health technologies, because we all are stakeholders. What we also are all learning at the same time, as we're implementing this, this is new and novel and technology we know is quickly advancing. Something that was new yesterday is already old tomorrow. And so, I think that's a really great example of why we need to collaborate together.

So I'm going to turn to Laurent. So Laurent, I know that you're able to see some of the questions as well. And there's been a lot of really great questions about what are the other diseases or disorders that you have started to look at and explore the use of Actimyo. And also, where do you think this is in terms of a lot of questions about the 95th centile stride velocity versus a six-minute walk? And I guess early thinkings about why that was kind of chosen to really be what you're going to be... You were looking at that data of a traditional six-minute walk with this, and how this offers an opportunity to maybe collect additional informative data over the six-minute walk.

Laurent Servais:

Yeah. So indeed, lots of great questions. And it's fantastic to see how the audience is engaged. But the stride velocity 95th centile is just one outcome. One outcome that was found to be probably, I'm not saying the best, but one of the best for Duchenne muscular dystrophy. Believe that disease in which you have a similar mechanism of proximal lower weakness leading to loss of ambulation, stride velocity 95th

centile can make sense. Like muscular dystrophy, limb-girdle muscular dystrophy. Then, of course, come other disease on which we are working, like multiple sclerosis, Angelman, Parkinson, ALS.

And in all these diseases, obviously, stride velocity 95th centile may or may not be the best outcome, because the key question when we develop a outcome in a disease is, what is important for the patient? What matters? What do we need to measure? And that is the role of the physician, to discuss with the engineer and say, "Well, look, an Angelman patient it's not exactly the same than the Duchenne. We are interested by the number of force, we are interested by the walking parameters. We all interested maybe by the stride velocity. And maybe stride velocity can be a candidate, but do you have any kind of measuring ataxia? How could we measure ataxia?"

And as you might imagine from multiple sclerosis, the same applies with similar questions. So we need to first evaluate what do we want to measure as a physician, and then to identify a potential outcome, and then to optimize this outcome, and then to validate this outcome. And that's a process that can be done, probably not disease per disease because our life is too short, but at least group of disease by group of disease.

So my personal opinion, but that's just my personal opinion, is that we could potentially put all diseases with lower proximal weakness leading to loss of ambulation, and to look at 95th centile stride velocity. And then we could put disease with ataxia, right? Like Friedreich's ataxia, maybe Angelman, some types of multiple sclerosis. And then we could put a disease with abnormal movement. But again, it has to start from the clinic and probably 95th centile stride velocity will not be the outcome for every single condition, and will be probably a very poor outcome in other diseases. So yeah, it's really disease dependent, and clinic dependent, because we have to start from the patient.

Michelle Campbell:

Thanks for that. And I think I really appreciate you talking about how it may not be the appropriate concept for different diseases, and that we really need to make sure we understand the disease, we understand the patient's needs, what's meaningful to them, and then what would make sense obviously in that clinical trial program we're trying to ask. So I really appreciate you talking about that aspect.

Leonard, I want to ask you, and I think maybe some of our audience members noticed that we do have our both CDER and CBER colleagues today, but one colleague group that is not here on this call that is extremely important to us is our colleagues in CDRH, our Center for Devices in the Digital Center of Excellence. And I was hoping you could speak to the audience about our collaborative efforts with our colleagues in that center, and how we work with them with the DHTs.

Leonard Sacks:

Sure, absolutely. We work very closely, both with the Digital Health Center of Excellence and other members of CDRH. And our collaboration sort of spans a range of different activities. First of all, I should probably mention guidances, the guidances that we put out on digital health technologies on Part 11 requirements for ensuring that the data from digital health technologies is reliable and so on and so forth, are developed very much in a collaborative fashion with CDRH and DHCoE.

Then I think folks on the meeting are probably aware that we've developed a framework for digital health technologies as part of our Prescription Drug User Fee Act commitments. And that framework has also been developed in collaboration with CDRH and with the Digital Health Center for Excellence. We have a steering committee within FDA that has a number of responsibilities, both to review or assist review divisions in looking at new applications that contain digital data and also to develop guidances and meetings and address issues of concern. And that steering committee has very significant representation from CDRH and DHCoE.

I think in reviewing applications, and perhaps my colleagues in the review division may add to this, the review is divided between many different sort of contributors, as Ami had said, and DHCoE and CDRH passed out pieces of the applications to review as they relate to sort of more the technological concerns. So I think in summary, we're sort of very much involved in collaborating with both DHCoE and CDRH at many different levels, really working very much together. We wouldn't manage without them.

Michelle Campbell:

Yeah, they are great and great resource for us and we're always constantly collaborating with them. Actually, the slides I presented today were from my colleagues in the Digital Center of Excellence for that.

And so we did get a lot of questions and we have a lot of great questions and I'm hopeful that we'll be taking these questions back to the agency and I think that will help us think about how to answer some of these questions in a public way through guidances or presentations, because you do have amazing questions from our audience members today. A couple questions, "I wanted to talk about the use of real world evidence and real world data and could that be considered?" And I would just encourage you to look at our guidances in real world evidence and real world data and what you would need to do for that to be useful in collecting the use of digital health technology information.

Additionally, we mentioned the multiple ways to how do you get come in and engage when it comes to use of digital health technologies. So first and foremost, you can put something into your IND program to get advice from your therapeutic review division. And then, as Leonard just said, we consult our internal experts and statistics from the Center of Devices when appropriate on the use of digital health technology, and so we do bring in our experts for that. So that's one avenue. We do have the qualification program which Laurent's talked about and mentioned their experience. And while it is not a requirement, it can come in and I do know that we've been actively looking at the work of this through that program. And so those are some of the mechanisms. And also we have some mechanisms through the DHT steering committee in CDER to be able to do some high level presentations as well.

And so I'm going to wind us down as we've really done a lot in an hour and I really appreciate that. And so I'm going to ask our panelist one final question and that would be what would be the one piece of advice that you would tell any stakeholder? So it could be a patient group, it could be an industry person, it could be an engineer, it could be a device developer, a technology developer. What one piece of advice would you give them that you think is critical to help advance the science and utility of digital health technologies in clinical trials to support our trial endpoints? So I'm going to start with my CBER colleague, Hussein, to see what your one piece of advice would be.

Hussein Ezzeldin:

Thank you so much, Michelle, and thank you for giving me the opportunity to start because this is going to be a hard one. I would say please come early as much as you can. I think that's probably the most advice I can give to anyone is that please come and share with us what are your thoughts and we'd be more than happy to engage with you, and as you mentioned, bring in the experts and help you navigate through all the challenges and come to a place where everyone wins.

Michelle Campbell:

Thank you for that. Ami, what would be your one piece of advice?

Ami Mankodi:

Yeah, thank you Michelle. So I would just say that be transparent, be collaborative, keep patient first. I think the need of our, as everyone knows, it's a standardization of the technology of the devices, and I think together we can make it happen.

Michelle Campbell:

Thank you, Ami, for that. Leonard, do you have a few final piece of advice to folks?

Leonard Sacks:

Well, I'm going to have to add it along on top of what everybody else has really said, but I'm going to come up with something different, which is really to focus on the most important medical needs. In terms of the right population, the diseases which aren't well served, the measurements which really aren't efficient in current drug development. I think focus on the medical need will get us further than just trying to advance the technologies in trials in general. And obviously I think in this rare disease environment we are already kind of well positioned there, but that would be my advice to focus on the medical need.

Michelle Campbell:

Thank you. Damien, for you as a developer and working on the technology side, what would be your piece of advice?

Damien Eggensteiner:

So I think the two obvious are collaboration and indeed trying to dig into this industry. But I think overall we need success stories, because I think it's been 20 years since the digital device or digital health tool fell short of their promises. And at some point we need to focus on medical needs, we need to gather everyone around it and we need to start the small but making success stories. And I'm sure that when you demonstrate the value of those digital endpoints, then you can start to navigate all the disease or the needs, like real life post and all of this, but we need to start small, trust each other, and really expand from there.

Michelle Campbell:

Thank you for that. And Laurent, I'm going to give you the final word on advice. I know you've provided a lot of advice and thoughts throughout this presentation being out there and really showing the early adoption of this and some of those challenges, but what would be that one piece of advice?

Laurent Servais:

I would say do not underestimate the amount of effort. This is something that you cannot do alone with your PhD student and a device out of the shelf. You need to have with you five or six different types of person, you need to have excellent engineers, excellent programmers, excellent statisticians. You need to have regulatory specialists, you need to have physicians, you need to have physiotherapist, and you need resources. So all of this is just impossible in one year or two years with a limited budget. I think that we had no idea about how long was the journey. I mean, it was successful as Damien said, but my best advice is if you want at one point to get an outcome that would be used in trial, that would be impactful in trial, this will need a lot of different people and a lot of different expertise that nobody has alone.

Michelle Campbell:

Well, thank you for that. I want to thank each and every one of our panelists today for being on here and really lean a very dynamic conversation in the use of digital health technologies. We knew we couldn't accomplish it all with digital health technologies, we know we have many more questions and answers, but hopefully everyone sees that, to advance the science and DHTs, and the utility and clinical trial endpoints, that we need to work together collaboratively, we need to think about the long-term goals and the medical unmet need, and we encourage folks to come early and often to the agency to get good advice.

And with that, I thank my panelists and I'm going to turn it over to my colleague, Mike Pacanowski, from the Office of Clinical Pharmacology, who is going to lead the next session on biomarkers. So I turn it over to you, Mike.

Michael Pacanowski:

That was fantastic. Thank you so much, Michelle. What a really great discussion in that session and some wise words to leave us with for further reflection. So good afternoon to you all from a hazy Washington, D.C. area.

Welcome to Session 2 of the workshop on Rare Disease Endpoint Advancement. In this session we're going to focus a bit more on developing biomarkers as surrogate endpoints. So as Michelle said, my name is Mike Pacanowski, I direct the Division of Translational and Precision Medicine in FDA's Office of Clinical Pharmacology. And we are very fortunate today to have here with us some key leaders with very diverse expertise who will share their views on the state of biomarkers as endpoints for drug development. I'm particularly looking forward to hearing about an instructive example from Patrick Nachman, which will be followed by panel discussion where we'll hear some additional industry and regulatory perspectives on development of biomarkers as endpoints.

Before we get into the featured case of our session, I'd first like to lay out a few bits of background information just so we're all on the same page and speaking the same language. If you could proceed to the next slide, please. So first I think most importantly is what is a biomarker? There's been a lot of efforts to define this very carefully. There's been an NIH and FDA collaborative working group that spent some time trying to carve out exactly what a biomarker is. And more or less, a biomarker is basically most measures other than how patients feel, function, or survive. It's really a defined characteristic that reflects normal biology, could reflect pathology, or it could reflect response to some external input.

Most commonly biomarkers are drawn from molecular, radiographic or other types of tests or assays. And that measurement is then turned into something that reflects what physiological processes or disease processes might be happening. And the purpose of doing this is manifold, there's a lot of reasons to assess biomarkers. It could be to evaluate one's susceptibility for developing a disease, it could potentially be used to diagnose a disease or a condition or predict response to treatment. And that's just a few of the many different potential uses.

What we're going to be focusing on in this session really are those biomarkers that are used to evaluate responses to therapeutic interventions. Next slide, please. Now, it's important to understand what the standard is for approving drugs. So for a drug to be approved for use in the United States, there has to be substantial evidence demonstrating that the drug is effective and the substantial evidence standards is met through adequate and well-controlled investigations. In most cases, two adequate and well-controlled investigations are needed, but in some case, a single adequate and well-controlled investigation may suffice, particularly if it's large and drawn from multiple centers in the United States or elsewhere abroad.

But also, we could have one adequate and well-controlled trial or investigation coupled with confirmatory evidence. And I'll say that confirmatory evidence itself also can take many forms ranging

from animal studies to pharmacodynamic responses among many other different lines of evidence. And it's this latter approach, this one adequate and well-controlled clinical investigation plus confirmatory evidence that's often the focus for developments of many drugs for rare diseases. Next slide, please.

Now, biomarkers come into play because an endpoint has to be developed for the adequate well-controlled investigation. And evidence that a drug affects how a patient feels or functions or survived is what we would normally be looking at to support traditional approval. And traditional approval can also be granted if we have a surrogate biomarker that reliably predicts a specific clinical benefit. So there's clear, robust evidence that shows that there is a relationship between a change in a biomarker and some clinical benefit.

What we see more commonly in the rare disease space is proposals to use biomarkers that are reasonably likely to predict clinical benefit as the basis for accelerated approval. And FDA has a great deal of flexibility in assessing whether a biomarker is reasonably likely to predict clinical benefit, and this allows the drug to be approved while we get evidence to confirm that the benefit exists in post-marketing trials. And the next slide, please.

Now, there's a number of different pathways that FDA can use to recognize biomarkers for a specific use. In many cases, a biomarker may be part of the standard of care that is established by way of community consensus. That is the case for many things that are commonly used in the clinic that are based on years of experience and large clinical studies. But otherwise, FDA has a couple of other pathways that the biomarker can come to be recognized. One is our ordinary regulatory approval processes where we're working with companies and designing clinical trials and their clinical development program, and there's a negotiated agreement on whether or not a biomarker is suitable to support approval.

Alternatively to that, there's the Biomarker Qualification Program, which is under the Drug Development Tools Qualification program. And this is a program where a biomarker may be qualified for use as a surrogate, for example, in a specific context that wouldn't need to be negotiated for each individual program. It's sort of broadly recognized as being useful for that purpose.

So that's just a little bit of background in terms of where biomarkers fit into the universe of FDA's regulatory activities. And with that little bit of background, I'd like to introduce Dr. Patrick Nachman, who will be presenting prior to our panel discussion. Dr. Nachman is the director of the Division of Nephrology and Hypertension at the University of Minnesota Medical School, and he'll be talking to us about developing proteinuria as this endpoint for IgA nephropathy. Thank you very much. And Dr. Nachman, I'll turn it over to you.

Patrick Nachman:

Thank you very much, Dr. Pacanowski, and thank you to the organizers for inviting me to be part of this panel and do this presentation. I am a little bit intimidated by Dr. Servais' beautiful talk and I'm afraid that mine is going to be a little boring in comparison, but what I would like to share with you a little bit of our experience in evaluating a potential surrogate endpoint in a disease called IgA nephropathy. Next slide. I just want to disclose that my team at the University of Minnesota has participated in clinical trials, some of which I will mention today, but I have no financial relationship with the clinical sponsors pertinent to this presentation. Next slide.

I will briefly introduce the Kidney Health Initiative for those on the call who are not familiar with it, I will also briefly introduce what IgA nephropathy is and review what our group has done in assessing protein excretion or proteinuria reduction as a surrogate endpoint for this disease, and close by addressing maybe some of our knowledge gaps and limitations and future directions. Next slide.

So first, what is the Kidney Health Initiative? Next slide. So about 10 years ago there was a memorandum of understanding between the American Society of Nephrology and our colleagues at the FDA to create a forum in which various stakeholders concerned about increasing and facilitating drug and device and treatment development in the nephrology world. And the stakeholders, as you can see in the slide, include our colleagues in FDA, NIDDK, professionals in the biopharma and biotech industry, manufacturers. Very importantly, patients and patient advocacy groups and their care partners are a big voice in the kidney health initiative, as well as physicians and professionals in academia. Next slide.

So about in 2016, we convened a group of colleagues across several institutions and levels of expertise, or not levels of expertise, but types of expertise, to explore surrogate endpoints in clinical trials for IgA nephropathy. Next slide. So for those on the call who are not nephrologists, I want to just remind everybody of why IgA nephropathy and what IgA nephropathy is. IgA nephropathy is the most common glomerular disease worldwide and it was first described about 55 years ago by Berger and Hinglais in Paris on a day when there was no strike in Paris. But even after 55 years and up until very recently, we did not have well-established therapies for this disease.

And part of the difficulty about really studying therapies and disease, if you look at the left-hand panel here, if you look at kidney survival or kidney function survival in IgA nephropathy, you can see that the survival is actually quite good. And even at 10 years or so, about 75% or 80% of patients will be with relatively acceptable kidney function preservation. So to do a clinical trial where the hard endpoint of end-stage kidney disease occurs this late really poses a challenge. Next slide.

The other challenge for IgA nephropathy is that there is a huge spectrum of severity of disease. There are some patients who have asymptomatic, relatively indolent disease on the green left-hand side of the spectrum. And then there are patients who are on the right-hand side of the spectrum where the disease is very rapidly progressive and maybe perhaps have reached a point where treatment may not be effective because too much damage has happened. Next click and the one after.

But there is a group in the middle where the risk of progression is identifiable, where an intervention is desired, but we did not have a strong surrogate endpoint to study and use in clinical trial. And the biomarkers, if I may use that term, that had been used to define this middle range where an intervention is desired, has been based on proteinuria largely. Whereas for the other end of the spectrum it's been based on creatinine or estimation of kidney function. Next slide.

So the unmet need was that we needed a better surrogate endpoint for use in clinical trials in IgA nephropathy to design trials that would allow us to test the potential efficacy of new therapies without waiting for the hard endpoints of having kidney function or end-stage kidney disease. So we convened through the Kidney Health Initiative, a multidisciplinary team including representatives from industry, regulatory, the FDA academics to discuss and determine candidates surrogate endpoints in IgA nephropathy. Next slide.

So we focused on proteinuria for the reasons that I mentioned before and went through the exercise of evaluating if proteinuria could be used as a surrogate endpoint and to what degree it could be used as a surrogate endpoint. So we evaluated the biologic plausibility, that proteinuria or protein excretion, abnormal protein excretion, could be on the causative pathway to loss of kidney function and IgA nephropathy. And there are a number of in vitro studies and to some degree more limited data that proteinuria can indeed contribute to the loss of kidney function over the years.

However, we recognized that there was actually limitation to the data, mostly that this association between proteinuria and loss of kidney function was not necessarily specific to IgA nephropathy. And also it has been recognized for a while that in IgA nephropathy the degree of proteinuria that can be associated with loss of kidney function seems to be much lower than in other diseases where there is excessive loss of protein in the urine. I just want to add also that one of the challenges in IgA

nephropathy is that we do not have a diagnostic marker, we do not have a blood test that tells us that a patient has IgA nephropathy, and we still rely on an invasive test to make a diagnosis, namely a kidney biopsy. Next slide.

So we looked at the data in support of proteinuria reduction as a surrogate endpoint from cohort studies. Next slide. And there is actually a wealth of cohort, prospective and retrospective cohort studies, that have shown that the degree of proteinuria at baseline is predictive of the long-term survival of kidney function. And you can see in the left-hand panel that patients who have more than three grams a day of protein excretion has a rather poor rapid decline in kidney function, rapid between quotation mark because you can see that the 50% survival did take up to five years, but in a graded way, patients with lower degree of proteinuria and almost close to normal degree of proteinuria still have some loss of kidney function.

Conversely, regardless of where the patient starts with respect to how severe their proteinuria is, if they can achieve a level of proteinuria that is less than one gram per day, the outcome of kidney survival is similar if the patient can get to less than a gram a day. Next slide and next click also. This slide also represents the same concept that baseline proteinuria is predictive of long-term outcome, but regardless of what kind of treatment is given, in this case this was two different types of treatment of renin-angiotensin blockers only or renin-angiotensin blockers plus corticosteroids. And at the bottom-right panel you can see that really the risk of progression depends on whether the patient can achieve proteinuria less than one gram a day versus more than one gram a day, regardless of how the patient gets there. Next slide.

So this was data coming from cohort studies and retrospective studies. What are the data coming from clinical trials in IgA nephropathy? Next slide. And for this part of the analysis, we really benefited from Lesley Inker and her group's analysis, really a meta-analysis of clinical trials that had been published at the time in IgA nephropathy. Next slide. And what this group did was a meta-regression analysis of... If you can just click three or four times, sorry. Of different interventions that had been tried for IgA nephropathy.

And without going into the details, you can see that if you look at what they have essentially analyzed is the effect of the intervention on urine protein excretion and whether there was a relationship between the effect on proteinuria versus the effect on the clinical endpoint, which was a composite endpoint reflective of loss of kidney function. Next slide.

While we were doing this, two new trials came up. One of them actually suggested that even though a drug could reduce proteinuria, that there was no associated demonstrably beneficial effect on kidney function, whereas another trial did show an association between reduction of proteinuria and a beneficial effect on kidney function. Next slide. So Dr. Inker and her group included these two new trials. Next slide, please. This is the testing trial, this is the IgA trial. Next slide. Next slide. And lo and behold, the relationship between proteinuria reduction and the effect on preserving kidney function could be maintained with these two new trials. Next slide. Next slide.

So the group really did feel that there was good evidence that proteinuria reduction could be used as a reasonably likely surrogate endpoint for the treatment effect on the progression to end-stage kidney disease in IgA nephropathy. Next slide. This work was published in 2019. Next slide. And very shortly after this concept has been applied to the design of two new clinical trials in IgA nephropathy. Next slide.

And indeed, earlier this year, these two trials have been completed and published at least the early phase of the trials utilizing proteinuria reduction as a reasonably likely surrogate endpoint. Both trials are ongoing in their confirmatory phase to really see if proteinuria reduction is indeed predictive of preservation of kidney function in the longer outcome. And conversely, the results of these two trials

will inform us on how well this reasonably likely surrogate endpoint predicts the clinical benefit in IgA nephropathy. Next slide.

So what are the limitations of our work and why did we focus on proteinuria? Next slide. Well, the short answer is not because proteinuria is a wonderful biomarker, but this is where our data was. Next slide. I think that in terms of knowledge gaps warranting future studies, first of all, the data is derived from what has been studied with the biomarker that has been studied and whether we can apply this better or to broader patient population and how best to apply this remains to be studied better. Next slide.

But I just want to mention that over the years, and this is a table that I put together probably 10 years ago, but it's still I believe relatively relevant, is that there has been a number of biomarkers that have been suggested for use in IgA nephropathy, but none of them has been well studied really for use in clinical trials. Next slide.

So I think that in terms of future direction, and this is a little bit my own personal opinion, I would love to see surrogate endpoint that are not just useful for clinical study design, but are also informative for the physician and the patient to know if they are responding to the treatment in a beneficial way. And proteinuria is really so variable from week to week, from day to day, with diet, that it's not a great biomarker of disease activity on an individual level, it may be great for use in clinical trial design.

Patrick Nachman:

I think that we need to do a better job at defining better, more specific markers of disease, of disease activity, and ideally, something that would tell us that the patient is in complete remission or that the disease is cured. How do we obtain and share data on specific biomarkers coming from various sources? And I just want to mention that in a different glomerular disease, namely membranous nephropathy, a new antibody test was identified about 12 years ago, antibodies to phospholipase A2 receptor. And earlier this year, the NephCure Foundation has convened a scientific workshop to help the community take this test, which is currently really validated as a diagnostic test, and see how we can analyze the use of this test possibly to monitor disease activity, disease response to treatment, and maybe perhaps incorporate that as a surrogate endpoint. Next slide.

So in summary, I think the exercise that we have done has helped us define a reasonably likely endpoint. We have been able to use this reasonably likely endpoint to bring two new therapies to market through the accelerated pathway. And conversely, these new trials will help us test the validity and the robustness of proteinuria reduction as a surrogate endpoint. And I think we should have more work to do to bring to fruition better and more disease specific and easier to use surrogate endpoints.

Last slide. I want to thank my colleagues, Aliza Thompson and our colleagues from really around the world who have helped us do this work, and really, the support of the Kidney Health Initiative in bringing this to fruition. Thank you.

Michael Pacanowski:

Great. Thank you, Patrick for that presentation. It was a really excellent illustration of taking available data, evaluating it, and turning it into something that was ultimately able to help inform regulatory decisions. So really appreciate you walking us through that case. We'll now move into our panel discussion. I'll ask my colleagues to join me on camera, and we'll do some quick introductions. So first, Dr. Nachman will be joining us on the panel providing some input from his perspective and kind of generating some of these data and analyses. We also have Lynley Thinner, who serves as the executive director of regulatory affairs at Travere Therapeutics, as well as Dr. Aliza Thompson, who serves as the deputy director of the division of cardiology and nephrology within CDER at the Food and Drug Administration.

So we had, I think, a really good case. And Aliza, you were very closely involved, I think, in the evolution of this biomarker throughout the various stages. But before getting into perhaps some technicalities related to that specific case, I'd like to ask you and Lynley to provide just some general reflections on biomarkers as endpoints, but also sort of what you see as some of the biggest challenges that exist in translating evidence about relationships with biomarkers and the disease process to something that can be used to support regulatory decision-making or drug development decisions in general. And Aliza, I'll begin with you, please.

Aliza Thompson:

Sure. So I think as Patrick noted in what I thought was an excellent presentation, though I may be a little bit biased, is that, at least in the space I work in, the kidney space, the outcome that we care often most about, which is progressing to kidney failure, is a late outcome. And so it's really challenging to design trials that look at effects on slowing the rate or essentially evaluate treatments for kidney diseases unless you're going to rely on a surrogate endpoint.

So at least in my space, really, we spend a lot of time thinking critically and really trying to understand the data supporting certain things that we measure in the blood or the urine or on imaging, and whether that can be relied upon to establish the effectiveness of a new therapy. In terms of challenges as relates to developing biomarkers as surrogate endpoints and what works and what doesn't work, I would say one of the challenges relates to sometimes even when there is data that can be used to critically evaluate whether certain biomarkers could be used as a surrogate endpoint, in fact, the people who have the data aren't amenable to sharing it. And I think the IgA nephropathy story is a beautiful story in that it was an example of many different groups coming together and really a willingness to share data to really advance the care and the ability to essentially develop drugs, therapies, to treat what is a very serious disease. So I think that's one issue.

And then I think the other issue is sometimes we simply just don't have the data. For some of our, and again, focusing on the space I work in and the kidney diseases, some of our diseases are very rare. It's really only recently that we've realized... We've taken some larger kidney disease and broken it down and realized that this is really a totally different beast here and a different disease with a new name. In those settings, it can also be challenging again because, yeah, the data aren't there.

But in either case, I think really, and I think this was highlighted in the earlier panel, the patient community can be a driver here. They can be a driver in terms of making sure that the various groups share data. Obviously it needs to be shared in appropriate way, protecting confidentiality. They're also a critical driver in terms of both funding and actually doing some of the natural history studies that are needed to really give us a good understanding of what biomarkers can and can't be used and relied upon as surrogates.

Michael Pacanowski:

Yeah, very wise comments. I think it is a necessity in a lot of diseases where the progression is slow to have these types of biomarker endpoints. And in rare diseases, you pointed out the common lack of data that I think is pervasive among the many thousands of rare diseases that exist. It just doesn't exist. And I'll pick your brain about that in a little bit with some additional questions. But Lynley, would like to hear from you a little bit about what you see as some of the big challenges facing drug developers with respect to trying to develop an evidence base that supports an argument that a biomarker could be suitable for approval decisions.

Lynley K. Thinnas:

Dr. Thompson did bring up an awful lot of the challenges we do face, and that's just the lack of data that's available for many of the rare diseases. Fortunately for IgAN, there was the consortiums that have come together to help to develop the data that we could then build upon and have a general sense that proteinuria would work as a surrogate endpoint that we could then test. It's one of the things that over time we have shown that some of the biomarkers that do get chosen may or may not lead to being able to show that earlier intervention can help in the long term. And this is really important to try to get to for an evidence-based decision, especially in cases like IgAN where these are, in some cases, slow progressing, and fast progressing as Dr. Nachman showed us in some of the examples. And sometimes it seems, while the patient communities are very active, they seem to not get as much attention. And recently this has been a good example that we should try to pull into other areas across other rare diseases and other divisions.

I think in addition, access to information. In our case we've had pretty good access with FDA and being able to have discussions early on. What are the appropriate endpoints? What changes do we need to make along the way? And we really did it with advice going through everything. And one of the things that I would recommend to those that are developing surrogate endpoints in addition to this pilot program that's being touted, there are a number of other pathways to try to use: fast track designation, breakthrough designation, the Type C biomarker meeting. These are all areas of where we can access FDA and be able to have a lot of really good discussion around what's reasonably likely.

I think one of the next steps is how do we move from reasonably likely to something that can be used for more traditional approval? How many examples do we need? And that's a challenge in looking at how do we then capture long-term data to be able to really help to confirm that a surrogate such as proteinuria could be used, either across glomerular diseases in general, or is it really more specific to one particular disease state? And these are some of the knowledge gaps that were alluded to that we do need to fill with respect to that.

And I think the other thing is just trying to get information more rapidly. While this was a good example where FDA, industry, academia, patient advocacy programs all kind of came together, the course of getting us to where we were able to gain a consensus that this was something that was reasonably likely to use as an endpoint in a study to be able to look at, took years. And so this isn't a process that is rapid. And it's trial and error, and there's lots of twists and turns in this journey to try to come up with what's best, trying to get therapies to patients who really need it the most.

Michael Pacanowski:

Yeah, that's very helpful. And do want to, I think, echo one comment that you made and just sort of reinforce the interest of FDA to try and advance therapeutics for patients with rare diseases. Establishing programs like the Rare Disease Endpoint Advancement Program is, I think, one step in that direction where we have continued engagement in a much closer fashion with the drug developers to get to a meaningful outcome that would ultimately hopefully pave away for a drug to be approved.

Dr. Nachman on the same theme, I was wondering if you could comment, I know you touched on this a little bit in some of your closing remarks for your presentation, about the need for some of these to translate to the clinic in a patient-centric monitoring kind of way. Are there any other more global challenges that you think need to be remedied in your perspective as a provider?

Patrick Nachman:

Sorry, trying to unmute. I think that there are a couple of things that maybe I could add to what Aliza and Lynley mentioned, but one thing that I've learned about what we have done is that I think we have more data than we think we have. And one of the challenges is that the data's not always acquired in

the uniform way or in a comparable way or in a poolable way. But if we do really go back and think about it and analyze it carefully and maybe do a little bit more of invest in bringing that data in a more uniform and comparable way, I think we can probably bring it to fruition. And we do have more data than we have.

I mean in terms of, if I'm following your lead in terms of how do we bring this from the clinical trial to the clinical practice, I think that this is where, and Aliza has reminded me of this more than once in our conversation, but simply because I feel that sometimes what works well in demonstrating that the therapy is beneficial at the level of the clinical trial is not always translatable to what the patient feels, functions, experiences and feels confident that they're getting better.

Earlier in the discussion there were mentions of clinical outcomes assessment of patient reported outcomes, and I think that this is another place where number one, FDA is helping a lot in developing these tools, and I'm very grateful for that, but it is certainly a challenge to try to develop those to bring them to the clinic, essentially.

Michael Pacanowski:

Thank you for that feedback. I had a number of questions. I think you touched on, I think, a few of the key issues around there's a need for collaboration, and we'll delve into that a little bit more, but I would like to touch briefly on the evidence and concept of clinical validation because it's come up a few times in the audience Q and A in various ways. Aliza, could you perhaps speak a little bit about reasonable likelihood, and from an evidence standpoint, what reasonably likely to predict clinical benefit means to you from your perspective, mainly with kidney diseases, but even more globally? Is it a matter of just understanding mechanism of the drug and how the disease progresses, or is it quantitative evidence such as we saw today or some mix of those things?

Aliza Thompson:

Right. And so that's really a challenging question because at the end of the day, it does sort of come down to a judgment call. But sometimes it's great. We have these opportunities to work with all of industry and to work with the patient community as well as trialists and to have that discussion as well over the level of evidence that's out there, and is this sufficient to support a conclusion of validated or reasonably likely? So I think there is room for discussion once you have these data. But just in terms of thinking a little bit about the level of evidence, we've never really defined criteria. There's no recipe for validating a surrogate or for what constitutes a reasonably likely surrogate. But I think we have sort of thrown out the buckets of data or some of the factors that we generally consider in deciding whether something you measure in the blood or urine or on imaging, a biomarker, can serve either as a reasonably likely surrogate, and support, therefore, accelerated approval or what we call a validated surrogate endpoint support traditional approval.

And one of the big buckets is sort of a mechanistic understanding. And by that I mean how well do we actually understand the disease process and the specific disease and really the role of biomarker in the causal pathway. There's some kidney diseases, and again, I'm going to always default to my space, where we really actually do understand it pretty well some of our genetic kidney diseases. There are others, and I'm going to say IGA nephropathy is perhaps more an example of this, where there could be actually be multiple pathways involved. And also the biomarker we've been talking about is actually... to the extent it's actually causing kidney damage is pretty unclear. But what we do think is that it's probably tracking with the injury to the kidney.

So I think that comes to another setting where we turned to other evidence and really needed a bit more then in terms of understanding the relationship between the biomarker and the outcome and

even wanted or had in that setting, I should be careful, some data from intervention trials, though none of which... Just want to highlight, all the intervention trials that Patrick showed you, none of them had led to an approval of therapies. It was sort of repurposing, right? Repurposing a number of studies that had been done that looked at a variety of different endpoints and trying to say, "Well, if I can get around... look at when they assessed proteinuria and can get some data there and also look at the later outcome, whether together as a whole, it looks like we're seeing this relationship hold."

Part of where I'm going with this, and I apologize if I'm going on a bit too much, but I think a critical factor here isn't just yes or no, it's a surrogate, or yes or no, it's a reasonably likely surrogate. I think a critical part of all of this is really the understanding of the quantitative relationship, which is what the work that's been done helped establish. And what I want to flag in this is really two aspects. One is there are kidney diseases where we could say, "Look, you take patients who have really high levels of proteinuria and you essentially bring them back to normal range." We'll buy it. We believe that you've essentially shut down the disease process, that you've done a good thing. There's both a lot of biologic plausibility as well as just really, really strong observational data that convince us.

And so that could serve as the basis for traditional approval. But lesser changes in this that we buy that you've done a good thing, that's really a setting in which we consider the surrogate or the biomarker as a basis for reasonably likely surrogate. So just want to highlight, it's not a yes or no. Some of this is tied to the magnitude of the treatment effect. And I don't know whether Lynley wants to speak to this further, but I think one of the things that the KHI project allowed was really, it can be challenging for drugs to totally shut off proteinuria, or in terms of if you look at the effect on the surrogate endpoint, to bring it back to normal range.

But I think one of the things that this project helped us do was understand the quantitative relationship so that we could actually be able to approve programs under the accelerated approval pathway where they maybe lower proteinuria by about 30% or something. So the proteinuria wasn't gone, but they'd had some incremental improvement. And it also allowed us to really have confidence in designing the confirmatory trials, understanding that based on this magnitude of a treatment effect on proteinuria, you understood how to design a trial that was likely to provide interpretable data on the ultimate benefit you're trying to verify. So just a few thoughts. I don't know, Lynley, what your thinking is.

Lynley K. Thinnas:

Yeah, and to build on that, many of these, especially in the IgAN studies looking at proteinuria, a lot of the other endpoints that we look at within the studies also help us to have and gain the knowledge that what effect proteinuria may be having on these other endpoints early on with interim analyses and looking at where things are trending. And then by the time you get to confirmatory, hopefully at that point you basically have showed, through the totality of your data, that you have everything going the right direction and/or you can confirm other things that go along with it.

So specifically if you're bringing someone back down into the normal range of proteinuria from a very high level, is that near a level for complete remission? And what's happening on some of your harder clinical outcomes time to ESKD and things of that nature, and seeing what that relationship is and helping to build that case as we put these forth to all the regulators to review, to make the case of the data and learning more about it.

Michael Pacanowski:

That's very helpful. So we are at time for this session, but I do want to very quickly, following in Michelle's footsteps, just to offer a final question. If you had a single wish to advance the field in five words or less, what would that be? And I'll start with you, Patrick.

Patrick Nachman:

So one thing that I would love to see, for example, taking the example of IgA nephropathy, is increasing the private-public partnership to help drug developers to actually invest and do mechanistic studies to learn about other biomarkers and how they could be used. This was more than five words, but I think that there's an opportunity there to help the developers help the science and vice versa.

Michael Pacanowski:

Very good. Lynley.

Lynley K. Thinnis:

In addition to that, I think it's also really important to maintain connections with the patient groups and advocacy to really understand how this affects the patient because there may be opportunities there for future ways of looking at the data to continue progressing treatments for this underserved population.

Michael Pacanowski:

Great. And Aliza, I'll let you be the final word.

Aliza Thompson:

Uh-huh.

Michael Pacanowski:

[inaudible 02:04:47].

Aliza Thompson:

Okay, well, I always like to say successful drug development takes a village and I think successful surrogate endpoint development takes a village. Got to put the patient and the patient advocacy group at the center of all this and we have to share data.

Michael Pacanowski:

Yeah. I think data sharing and collaboration and all of those elements are extremely critical. So we are at time, a couple minutes over. I would like to thank you, Patrick, Aliza, and Lynley for really engaging in a productive session. We're going to take a 15-minute break. We'll be returning at 3:20 Eastern Time. And Naomi Knoble will be walking us through a session on clinical outcome assessments. I want to thank the panelists and the organizers for a wonderful session, and look forward to hearing the rest of the workshop. Thank you.

Patrick Nachman:

Thank you.

Naomi Knoble:

So hello everyone, and welcome back from the break. My name is Naomi Knoble and I am a pediatric psychologist and the associate director with an emphasis on rare disease measurement science in the division of Clinical Outcome Assessment within the Center for Drug Evaluation Research at the FDA. And I would like to warmly welcome you to our third session of the day, entitled Considerations in

Developing Rare Disease Endpoints: Clinical Outcome Assessments. So in this session we'll focus on an explored innovative example here of an endpoint that includes a COA and you'll learn about relevant endpoint development, challenges, opportunities, and possibly some solutions as well. This session will end with a panel discussion where the panelists will share some thoughts and opportunities and challenges in developing rare disease endpoint using a clinical outcome assessment, or a COA.

And before we turn to Dr. David Rousso's presentation, just wanted to briefly highlight our Patient-Focused Drug Development Guidance Series. So next slide. Thank you. These four guidances are actually reflective of all three of our centers, Center for Biologics, Drugs, and also Devices and Radiographical Health. And Guidances 3 and 4 in this series are currently in draft version, and we'll move to publication after we've integrated public comments. And they specifically address clinical outcome assessment development, modification, selection, and ultimately, use in clinical trial efficacy endpoints.

So clinical outcome assessments measure how patients feel, function and survive. And FDA broadly defines four types, which are patient reported, observer reported, clinician reported, and performance-based outcomes. Our first session today presented an innovative performance-based outcome that leverages digital health technology. And in session two on biomarkers, Dr. Nachman and others noted the important role of clinical outcome assessments in clinical trials as new biological indicators don't always reflect changes that are noticeable to patients. And so COAs can be critical to the success of interpreting clinical benefit from clinical trials. And Dr. David Rousso will offer another innovative example here of the use of a clinical outcome assessment. And so it is my honor and privilege to introduce Dr. David Rousso. He is the US Medical Affairs Therapeutic Lead for Ophthalmology at Spark Therapeutics. And so thank you, Dr. Rousso. I'll turn it over to you.

David Rousso:

Thank you, Naomi. And thank you to the organizers for inviting me to give this talk today. So yes, my name is David Rousso. I'm the therapeutic area lead for ophthalmology at Spark Therapeutics, and I'll be speaking today on the multi-luminance mobility test as a novel clinical outcome assessment that was developed for the phase three clinical trials for Luxturna. And as many of you may be familiar, Luxturna was the first FDA-approved gene therapy for a genetic disease for patients that are affected by a rare form of inherited retinal dystrophy due to biallelic mutations in the RPE65 gene. Go ahead and click to the next slide.

But before getting into the details of the MLMT, I thought it would be helpful to provide just a quick refresher on how the visual system works normally and how RPE65 mutations cause loss of vision in a very specific way, which will help us to understand why the MLMT was so crucial as the primary endpoint for this clinical trial. So normally, vision works when light from the outside world is focused by the lens onto the retina, which is located to the very back part of the eye. Go ahead and click to the next animation. And within the retina specialized cells called photoreceptors. Click again. And photoreceptors essentially transform this light signal into electrical signals through a process called the visual cycle. And this ultimately gives rise to our perception of the visual world around us.

Now importantly, there are two types of photoreceptors, the rods and the cones. And these are important for giving rise to different aspects of vision. So the cone cells are located in the very central part of the retina, and these give rise to our high acuity, high detail central vision. And the rod photoreceptors are located throughout the rest of the retina. And these mediate our peripheral vision. They're also more sensitive to light and therefore mediate vision under low light conditions such as at night. Next slide.

So RPE65 is a gene that is critical for the visual cycle to work. Click the animation. So in the... Oh, back. There we go. Yeah. So in the absence of RPE65 gene function, the visual cycle is no longer able to

function correctly. The photopigment is not able to regenerate itself, and this leads to a loss of visual signal and loss of vision in the patient. The goal of gene therapy here is to restore RPE65 gene function back within the retina. Next slide now.

So although RPE65 gene function is important for both rods and cones, it is the rod cells that are typically the earliest and most severely affected in this disease. And so patients typically present with symptoms that are indicative of a loss of rod-mediated vision. So these include symptoms like decreased light sensitivity, diminished visual field, or diminished peripheral field, nyctalopia, which is also known as night blindness, nystagmus and poor adaptation to low light. There's also progressive loss of central detail vision as well. But this clinical picture is very different from some of the more common retinal diseases like age-related macular degeneration, which primarily affect central vision. Here in this disease we see primarily loss of rod-mediated vision. And so these symptoms really increasingly limit an affected individual's ability to independently navigate their environment.

David Rousso:

We limit an affected individual's ability to independently navigate their environment, especially under low light conditions. Next slide. So historically there have been a number of different visual function assessments that have been used to measure vision in these patients with these types of inherited retinal diseases. These include visual field-testing to measure peripheral vision, full field light sensitivity threshold testing, FST, which is a measure of overall light sensitivity within the retina. Even electroretinograms can measure the electrical activity in the photoreceptors and response to light. And of course visual acuity is the primary measure of our central vision, but none of these assessments alone really captures the full range of visual impairments in patients that are affected by this disease due to bio mutations in the RPE65 gene. And so there was a need for a novel assessment to measure these types of visual functions in a clinically meaningful way. Next slide.

And so to do this Spark and study investigators really sought to go beyond these conventional measures of visual function, which essentially measure how well the eyes are functioning and instead develop an assessment of the overall integration of these different aspects of vision into a single measure of functional vision, which describes how well the person is able to function in going about activities of daily living, such as mobility and navigation for example. Next slide. So this led to the development of the MLMT as a novel clinical outcome assessment to measure functional ambulatory vision in patients across a range of different lighting conditions. The test was initially developed at the Children's Hospital of Philadelphia as an exploratory endpoint in the phase one two clinical trials for Luxturna. And in these early trials, some of the findings showed that patients who received gene therapy were able to go through a mobility course more quickly and with fewer errors after receiving treatment.

And so this led to the development of the more advanced version of this test, the Multi-Luminance Mobility Test, the MLMT, in preparation for the phase three clinical trial. And this was ultimately used as the primary endpoint and importantly, this included input from the FDA as part of the development process of this test. And that helped to establish the high degree of rigor in terms of the reproducibility and reliability of this test. And I'll get more into that in a moment. But essentially the test was designed to provide a clinically meaningful assessment of functional vision and to evaluate potential changes in functional vision over time, including after intervention and again intended to measure functional ambulatory vision at light levels that are encountered in activities of daily living. Next slide.

And so this is what is shown here. This is the MLMT as it was used in the phase three clinical trial. Essentially testing the patient's ability to follow visual cues while avoiding obstacles over diminishing levels of ambient light, ranging from very bright 400 lux level, which is typical to a working office environment all the way down to one lux, which is very dim akin to stepping outside on a moonless

summer night and various light levels in between at roughly half lock steps, which really capture different levels of lighting conditions that one encounters in going about activities of daily living. And essentially what the test is measuring in terms of the outcome of the measure is the lowest light level at which the patient is able to complete the course. Next slide.

So as part of the development of the MLMT, as I mentioned, there were several important standards that were established to ensure testing rigor and reproducibility. And these included 40 minutes of dark adaptation for every subject prior to testing, testing each eye separately in case one eye was dominant as well as both eyes together bilaterally. There were 12 differently standardized courses that were similar in difficulty in terms of the number of turns and number of obstacles, but helped to mitigate against potential learning effects that might otherwise occur. And performances were videotaped and recorded and coded by external graders that were independent and massed to the various study and conditions of testing.

A rubric of scoring the performance of the test was based on speed and accuracy of performance and again, resulted in a final score based on the lowest level of light that the patient was able to successfully complete the test. I should also mention here that there were several other standards with respect to the testing sites themselves. So the testing sites also had construction of a dedicated room with lighting systems installed overhead and light levels that were measured daily using calibrated light meters at five different points throughout the course and initiation to any testing runs. Next slide.

So we assessed the validation, we assessed the validity of the MLMT as part of a validation study to assess the construct and content validity of this test. This test included 54 children and adults, 26 were normal sighted individuals and 28 were low vision due to various forms of inherited retinal disease. And among the normal sighted patients in this validation study, essentially all normal sighted individuals showed consistency between their performance on the course between baseline and after one year of follow up. And all of the subjects were able to pass the course at the lowest light level of 1 lux. However, among patients within the IRD group, there was a wide range of baseline visual function and functional vision within this group and there were no improvements seen over one year. And in fact, eight subjects actually declined in their performance over the course of one year by as much as one or two light levels.

Next slide. So some of the key findings from the validation study, again, as I mentioned, there were 12 differently configured courses that were designed and randomized between each test to help mitigate against learning and we showed that they were similar. In fact, in terms of their difficulty between each test, the scoring system was found to be highly reproducible with good inter and intra greater reliability and also subject test and retest reproducibility. And importantly, we also found that the MLMT accuracy scores showed a relationship with other measures of visual function like visual acuity and visual field. So there was correspondence with other measures of vision as would be expected. The test could also distinguish between the vision impaired and normal sighted subjects. It could identify a range of different levels of functional vision within the low vision group and could detect changes in performance of functional vision over time. And so based on these findings from the validation study, this supported the use of the MLMT as a clinical endpoint and therefore its use as the primary endpoint in the pivotal phase three clinical trial for Luxturna. Next slide.

And so just in the last few minutes of my presentation, I want to touch on some of the important learnings from the use of the MLMT as a measure of change in functional vision from this study. So this was a randomized controlled study that included an intervention arm as well as a control arm. The control arm did not receive treatment but was monitored for one year and then given the opportunity to cross over and receive treatment after that one year period. The MLMT bilateral performance was selected as the primary efficacy endpoint, but there were also several secondary measures that included

some of the more conventional measures of visual function that were also incorporated into the test or into the study design. Next slide.

So here are the MLMT data at one year showing lux level change scores for the bilateral MLMT as well as only using the first treated eye, which was the patients worse seeing eye and showing a median change in light level performance on the MLMT of two in the intervention group and in the control group, no change was seen over the control period of observation. Next slide. So to give you a better idea of what a two light level change on the MLMT test actually looks like in terms of a change in functional vision, I want to play this video that depicts a representative subject performing the MLMT at baseline at one lux resulting in a fail, and then the same subject performing the MLMT at one year post-treatment at the same light level of one lux, but then passing at that point, and go ahead and play the video.

So you can see that after intervention improvements in speed and accuracy in the performance of the test at one year post intervention compared to baseline. And this represents a two light level change in mobility testing on the MLMT, which was the median change for the intervention group. Next slide. So shown here is a swim plot of the individual patient level data for all the subjects that were treated in the phase three trial. And this data is important because it reveals one of the unexpected findings from the MLMT data from this trial and a potential limitation of this test. And this is that quite a few individuals in the intervention group, which is shown on the left in blue, showed improvements on the MLMT down to one lux, which was the lowest light level tested. And so this represented the maximum change that could be measured on this test.

We did not design the MLMT to test mobility at lower light levels lower than one lux. And this was also seen on the right in control group, which are shown in black over the control period, but then in green after crossing over and receiving intervention eight out of nine subjects in that crossover group performing on the MLMT at the lowest light level of one lux. So overall there was a ceiling effect in this measure, which could potentially limit the test ability to show changes beyond one lux for some patients. And so this leads us to the next slide in regards to the importance of having still good secondary measures that are, may even include conventional measures. As I mentioned, we had several other secondary measures including the full field light sensitivity threshold test, which is just a measure of retinal light and sensitivity overall.

Next slide, and shown here is a scatter plot between the MLMT and the FST data at one year. The green dots represent the intervention group and the blue dots are the control group. And note that the FST on the Y axis is on the log scale. And while we see here, there was good correlation between the FST and MLMT. The FST test has a much wider dynamic range in terms of the magnitude of change that could be measured. And so we, by pairing the MLMT and the FST data, this helped to capture further changes in light sensitivity in patients whose changes in functional vision as seen on the MLMT could have been limited due to the ceiling effects of this test. And so this pairing of the FST one of the secondary measures with the primary measure of the MLMT really helps to contribute to the overall totality of evidence from this trial. And in the interest of time, I will turn it back over to Naomi. That was my last slide. Thank you.

Naomi Knoble:

David. Thank you so much. That was a terrific and thorough presentation and we're all very grateful. There are so many pieces of the MLMT's development and use that are just a great illustration of good COA development and I appreciate you pointing out there is this ceiling effect. So while that might not be ideal, it wasn't a deal breaker and it's great illustration of challenges but also solutions. I love how you highlighted the role of the totality of evidence, those good secondary assessments to support your primary efficacy endpoint. Oh gosh, there's so much that went into that. The clearly identified aspect of

vision that you measured, how carefully that was crafted, functional vision, the rigorous reader training and adjudication, the validity assessment. There are more pieces, and I won't bore you with my comments here, but let's turn next to our wonderful panel discussion here.

And I'd like to invite our panelists to turn their cameras on as I introduce our group here. I am just honored and I think it's lovely for all of us to be joined by Abigail Luo, who is a mathematical statistician in the division of biostats on the Office of Biostatistics and Pharmacovigilance within CBER, or Center for Biologics at the FDA. We're also joined by Dr. Lindsey Murray, who is the executive director of the Rare Disease Clinical Outcome Assessment Consortium at the Critical Path Institute. Dr. Lei Xu is the chief of general medicine branch two in FDA's division of Clinical Evaluation of General Medicine in the Office of Therapeutic Products, also with our Center for Biologics, CBER. And finally of course, Dr. David Rousso.

And so thank you all so much for being a part of this panel together and I'd like to move here into our discussion portion of the session and briefly just turn to our list of questions that we have for the session. But of course we'll also be taking questions from audience Q&A as well. And so I'd like to start us all off here. Rare disease is a challenging space. We all know this, but we've also seen, I think David's presentation is a great example. There are terrific and innovative solutions that can be brought to bear on rare disease measurement and successful trials. So what are some of the biggest challenges and solutions that stakeholders experience in developing and using clinical outcome assessments for rare disease research? So David, why don't we turn to you since you have just given us such a nice example of some challenges and solutions and then I'll let our other colleagues join in.

David Rousso:

Yeah, sure. I mean certainly for rare disease, one of the challenges is just due to the nature of a rareness of the disease. And so there can be challenges and limitations in terms of the understanding of the disease's natural history. We've heard others refer to this challenge, which kind of pervades the entire area of our disease. So there can be little known in terms of the variability of symptoms, age of onset progression. And this can make it very difficult to develop endpoints that accurately reflect the disease and to demonstrate potentially the effects of potential treatments. In the case of Luxturna, one of the ways that we overcame this challenge was by conducting our own multi-center natural history study. This was done in parallel to the clinical development program for Luxturna. And by doing this, we were able to characterize changes across a range of different measure visual measures over time within this specific disease population. And of course this also helped us to better understand the significance of the data that we were seeing from our phase three clinical trial.

Naomi Knoble:

Wow. And David, just a follow up question with touching on the understanding the disease's natural history, I think you're right to point out that's a huge challenge in the rare disease space. And what would you say to stakeholders who often say, why spend the time on a natural history study, why not just jump in to drug development or product development?

David Rousso:

Well, I think that would presume that conventional measures of clinical measures would be relevant to a rare disease population that was perhaps very little understood and there would need to be some investment upfront to understand what kinds of clinical features of the disease may change over time and what's more meaningful to patients in terms of what they care about for potential outcomes after a potential intervention. So I think that's just really, really fundamental to lay a foundation in

understanding the disease that you're aiming to intervene and treat to develop what the proper endpoints would be for any potential studies.

Naomi Knoble:

Yeah, thank you so much. I think speaking for my FDA colleagues around the panel, certainly what we see often is what works clinically for clinical practice often fails in clinical trials or can fail in clinical trials. And really appreciate your point about focusing on what's important to patients and integrating that into outcomes as well. And Lei, I'd love to turn to you, what from your perspective, are some of the challenges and solutions in developing COAs and rare disease research and trials?

Lei Xu:

Hi, good afternoon everybody. Thank you for inviting me to be part of the panel and also thank you Dr. Rousso for the very nice presentation. And that was the first gene therapy BLA I was involved in the review process. So I just want to echo what Dr. Rousso just said. It's really important to try to at least have some understanding of the natural history. I mean, for the rare disease, a lot of times we don't know what the natural history is. For example, for Luxturna, I remember the phase one study, the efficacy assessment was visual acuity, very conventionally used endpoint for the ophthalmology indications. I don't think that endpoint really captured what the product could do, but fortunately they do have the natural history data and they also had that, I think the early version of MLMT in place during the phase one study.

So that's when that this particular COA was further refined and so that it could be used in the phase three study. And I remember during that period, we have a lot of back and forth regarding whether using one eye or both eyes or I mean how many points of change it could be clinically meaningful. And we do have some disagreement, but I mean fortunately the Luxturna, I mean, does have a relatively big treatment effect. So even though with those disagreement we were able to come to the mutual agreement that this is a clinically meaningful outcome with this treatment.

And so I want to emphasize it that when the sponsors out there are trying to develop some clinical outcome assessments for their conditions that they want to study, I think it's really critical to understand the disease first. And then when they are trying to build up those items in the COAs, they should really think carefully about the clinical meaningfulness of those items because sometimes we do see something related to the physical exam findings, which is important sign, but it may not be clinically meaningful to those patients, which as you mentioned earlier, reflects how patients feel, function, or survive.

And also I really encourage sponsors to start the conversation with us early in the process and we have worked very closely with the COA team in FDA to get them involved so they could provide constructive input from the COA perspective and we could provide input from clinical perspective.

Naomi Knoble:

Oh, go ahead. Sorry.

Lei Xu:

I think one thing, because I think over the last several years we have approved several gene and cell therapy products. I think all of them were for rare conditions. And some of the products, especially some of the AAV-based therapy products, which giving systemically it is like a one-time administration. So we really hope the product could bring some benefit to patients who receive them because after that they at least with available knowledge, we have not had a way to re-administer a similar product of a

AAV product. So I think that's really important to give something that could potentially be effective and safe.

Naomi Knoble:

Yes, thank you so much. And there's so many great takeaways from your comments here. I think what you've ended with here about how frankly in the gene therapy space, especially the stakes are so high because it is really one chance, and frankly, those patients aren't necessarily eligible for their other clinical trials either. And so the risk, there's such massive unmet need for treatments in the rare disease space broadly, and the stakes are seemingly so high specifically for these biologics and gene therapies. And you highlighted some other great points too that I think were echoed in the first session that we had for today's workshop as well, just the road in the pathway to developing the MLMT.

It wasn't quick necessarily. I mean David, you could speak to that more, but it does sound like there was an iterative process. And again, just highlighting how that very traditional clinical measurements really wasn't quite right for this particular indication in the context of use. Great points you've made. Lindsay Murray, I'm wondering if there's anything you'd like to add here just to this question of challenges and solutions, especially for COAs and rare disease.

Lindsay Murray:

Well, I have to say, David, I love your presentation. I think the functional aspect, linking the COA to the real functional aspect is so critical. In rare disease we see the heterogeneity between and within patients is just so extreme in a lot of these diseases that it's really hard to pick one test that's going to be applicable to the entire population. And so I think taking it a step further and looking at the function is so crucial. It's great if a child can tap their fingers more, but what does that actually mean in their daily lives and how do we apply that? But it really also allows us to address some of that heterogeneity because there might be multiple symptoms being experienced, but you can measure the functional impacts are similar across patients even if the direct symptoms aren't. So I think that's a real opportunity to better maximize the data that the patients are giving us.

And I really can only echo the importance of natural history data. And I know in very small rare diseases that's just not always going to be feasible. But having, I think even a minimum of real depth of qualitative data that you collected from patients to really talk with them about what are their key symptoms and complaints and what's the burden of this disease for them, even if you can't measure that quantitatively over time, I think really can help direct us in where we need to be aiming for, what's important to the patients, what are the things they'd like to see improved with treatment? And so I think that's even in cases where we can't have those robust natural history studies, make sure you're talking to your patients, make sure what you're treating is important to them, that that's really going to make a meaningful impact on their lives. Especially if you're doing these drug trials where they're going to have gene therapies and cell therapies where they have one shot at it, better make sure that that shot's on goal, they're actually targeting something that people care about. So...

Naomi Knoble:

Yeah. Lindsay, thank you so much. And I really appreciate the point you've echoed here that is made by Lei as well and of course by David, just what is clinically meaningful benefit, what is a clinical benefit that we can clearly see? And when something is functional and we understand how to translate that into how a patient is functioning in their daily life, it supports all of us with that interpretation of what is meaningful change. Absolutely. Abigail, I'm not going to leave you out of the conversation, but I'm going to pivot to another question here in regards to how from all of your perspectives, can stakeholders,

including patients and advocates work together to advance the use of clinical outcome assessments in rare disease product development? So I'll turn to you first.

Yuqun "Abigail" Luo:

Yeah, thank you, Naomi. Thank you. It's all very well said. David's presentation. And Lindsay and Lei and Naomi, you have already provided a lot of input. So I would actually touch on three points. So I'm a statistical reviewer at FDA, so maybe it's my vantage point is a little bit different, but where all contributing to this rare disease endpoints. So I think the first point I would like to highlight, and Naomi, you already mentioned that FDA has developed and is continuing to develop resources to incorporate the patient's voices in medical product development and regulatory decision making. And one of those, so these include perspectives from patients with rare diseases. And Naomi, you mentioned that you highlighted one such effort in relationship to endpoints based on clinical outcome assessment. And that's the series of four PFDD methodology guidance documents.

I would especially like to highlight guidance four title, Incorporating Clinical Outcome Assessment into Endpoints for Regulatory Decision Making. And that's very relevant to our topic today. And the draft of this guidance was recently released for public comment and the commenting period will close on July 5th. So we still have about a month. So I would encourage every stakeholder to submit your comments to us to help us improve the guidance, to incorporate your insights. And another point I would like to highlight is that I think I have heard a lot of comments about this is really all hands on deck everyone. So there should be a lot of collaborations across stakeholders. And one thing I would like to highlight from my personal experience is that this, were also in role collaboration between different disciplines. So in my role in FDA as a statistical reviewer, so we actually review a wide range of indications like Lei mentioned, that we have a lot of activities now due to the advancements that we have now activity on gene cell therapy that are in late phase development.

And we have quite a number of approvals in the past few years being one of the first. And so we have a lot of indications. So as a statistician, anytime I get a indication, I actually went to read all the context like the disease and the mechanism of the product and the current treatment landscape. And so that gives me a context like to make my statistical expertise relevant to my clinical colleagues and conversely colleagues from other disciplines will also expand their understanding of statistics to the extent that we can have effective collaboration. And these collaborations are very important when we can communicate and collaborate effectively, not only in endpoint development but also in aligning the clinical questions, clinical objectives, and corresponding with the statistical analysis and result into presentation. So I don't know whether I still have some time, I have another point. Do we still have time?

Naomi Knoble:

Yes. Although if you don't mind, I'm going to pivot to David and then we'll come back to you, yeah. So David, I'm curious from your perspective as a sponsor, how do you see stakeholders including patients and advocates working together to advance COAs in rare disease development?

David Rousso:

Sure. I think what I see the big potential for is just in collaborations to not only develop but also expand the use of COAs that have already been established for different rare diseases, which especially, which may share certain similarities in terms of the disease pathophysiology and also the way in which the disease impacts the activities of daily living that really ultimately matter for the patients. So for example, just based on what I've presented earlier, there are other rare genetic forms of vision loss that also

affect a patient's ability to navigate, especially under low light. And Spark has been working with other companies to share learnings from the development of the MLMT and potentially the possibility for licensing its use for other clinical studies as an endpoint for functional vision and ambulatory navigational vision to build upon the work that has already been done and just standardizing this measure of functional vision or there has really been a lot of investments already made.

Naomi Knoble:

Yeah, I appreciate both Abigail and David, both of you highlighting this just the critical role of collaboration and that I think that's been part of the theme of today even as well. And Lei, what you highlighted earlier, which I think is, if we had a motto at FDA, it would be come speak to us early and often Lei is, are there any other points you'd like to tack onto this?

Lei Xu:

No, I don't. I mean, I don't know. I saw some questions and Q&As here.

Naomi Knoble:

Yes.

Lei Xu:

In the QA box. I don't know whether...

Naomi Knoble:

Yeah, we'll turn to those in about a moment. Yeah, I just wanted to touch base with you briefly here. A question for you actually, Lindsay, is I'm curious how can sponsors identify existing COAs that could be reused or modified for new applications?

Lindsey Murray:

Well, we are the rare disease clinical outcome assessment consortium is working on that problem. Naomi, you're very aware and we are actually about to launch later this month. We'll launch our rare disease COA resource, which we've actually looked at pediatric non-oncologic patient populations and for the domains of self-care, gross motor function, fine motor function and communication and language. We've done that. We've gone through the existing measures, we've evaluated what's there against regulatory expectations from both FDA and EMA and then had internal expert panels and external...

Lindsey Murray:

And then had internal expert panels and external advisory panels review the results and really come together to say, "Look, these are the best tools that measure these domains of interest." And the domains were really around those ADLs because that's what that's impacting patients' lives and caregivers' lives. And so, we've done that heavy lift, and there's a short list now of here are the 10 or 15 best communication and language measurement tools that are currently available. And sponsors are still going to have to go in and look at the specific context of use that they're applying it to, but there are resources available to help with this and the resource is continually being built on. We're working in domains of pain and sleep now, expanding it, and so this work is ongoing. There are also other great

resources online for people who have collected core outcome sets for specific rare diseases. So, there are places for people to go.

Naomi Knoble:

Lindsay-

Lindsey Murray:

We're very excited to be ready to launch this rare disease resource.

Naomi Knoble:

Yeah, this resource has been a long time in the making. And so, it's the rare disease, (COA) resource, from the rare Disease (COA) consortium through C-Path, which Lindsey heads. I'm the CDER liaison, and we have CBER representation there, as well. And I think it's a reflection, too, of FDA's commitment to accelerating rare disease medical product development, as well. And it's been a dynamic partnership, right? We have our clinical experts, we have academic subject matter experts, industry members, NIH colleagues, and FDA, of course, contributed, as well. And I really appreciate what you highlighted here that it's an essential starting point for that conversation of how do you select a clinical outcome assessment for use in a clinical trial? And FDA was involved, and of course it's not an endorsement, but it is a launch point. And I'm with you. It's really exciting to have this launch, and I think, ultimately, it'll have a positive impact for people living with rare diseases. And they are always the number one stakeholder for everything that we're doing. Yeah. Thank You.

Lindsey Murray:

Yeah, I think it really gets back to this collaboration, as well, and how do people engage. C-Path, we have 11 rare disease consortium or initiatives right now that are made up of patient advocacy groups, sponsor representatives, FDA, academia. There's a lot of opportunities there. It's a unique opportunity within C-Path to have all of those people come together in one place and have this pre-competitive collaboration.

Naomi Knoble:

Yes, I totally agree. Turning to a question that we had from the audience, and, Abigail, remembering that there was another point you wanted to make, I wonder if you could take this one first for us. And the question is, how do you decide or modify an endpoint if the patient enrollment is slow? So, maybe a patient per month. And then, of course, that creates a challenging circumstance to reach the minimum number of patients required. As a statistician, do you feel like you can take that first? Or do you want me to punt it over to David to speak to slow enrollment?

Yuqun "Abigail" Luo:

Yeah, I think I do sympathize with when you are dealing with your disease it may be challenging to reach the sample size that you initially plan expediently. So I think it's a conversation that was to have with FDA very early on and explore different scenarios. So I think for rare disease, usually people have a good sense about the prevalence, et cetera. You usually have several big centers that are specialized in those patients. So I think so, and I think David mentioned, right, earlier on that there are a lot of work that need to be done and when you launch into that phase three study, so a lot of this work need to be done before you went into phase three. And then if you do encounter problems, I think not necessarily in the

realms of the disease, but sometimes the pandemic disruption, so like earthquake or something, okay, maybe this fire in Canada that's polluting our air right now.

So there are a lot of things that can influence how trial is ongoing. And then sometimes the sponsor may want to change the conduct of the trial, that's the time to talk to FDA, but I would urge the sponsor to at the planning stage that really, really spend a lot of effort at the front end including finding the right dose. They probably can speak to that as well. So I think sometimes there's this desire to get effective treatment to patients fast, but also we do need to try to make our effort up front. David, do you have anything to add or anyone else?

David Rousso:

I would only add, again, just kind of the theme of collaboration. Our phase three involved two study sites and enrolled 31 subjects, but for the natural history study, there were a certainly larger group, 70 subjects for that. And that involved seven centers around the world. There were three continents and seven centers around the world. So I think just collaborating with other centers even internationally if needs be to ensure that you're able to incorporate enough data from patients to characterize disease or potentially measure in intervention effects.

Naomi Knoble:

Yeah, thank you. The emphasis on global collaboration is so critical and Lei, I'm curious if you have any additional comments here as well.

Lei Xu:

Yeah, I think just one quick point Abigail just mentioned. Talk to us if you really encounter those issues and also go, I mean go back, to your enrollment criteria. I mean, do you really need that many exclusion criteria? Which could limit the enrollment. I think that's the only thing I want to add. Thank you.

Naomi Knoble:

Thank you so much Lei. And David, a question for you and a question came in from the audience, what types of other measures were used to establish forms of validity, like convergent validity? Did you use PROs, ClinROs, PerfROs? Can you just briefly give us a summary of that?

David Rousso:

Yeah, there we looked at other measures of visual function that I alluded to already. There was also a visual function questionnaire that was utilized, but it was not custom for IRD, but it helps to support the validity of the MLMT and that was published as part of the validation study as well.

Naomi Knoble:

Yeah. Thank you so much. And we have just a couple of minutes left here in our session and I want to ask all four of you my very favorite question, which is what gives you hope for rare disease medical product development and Lei? We'll start with you. Lei, I think you're on mute.

Lei Xu:

Oh, I'm sorry. So I think the hope I have for the rare disease comes from the experience our office has been experiencing over the last several years. I remember when I joined the agency about 13, 14 years ago, and most of the programs were in the very early phase of clinical development. We may get one

BLA every few years, but I think over the last six, seven years, the number of the BLA submissions has been increasing substantially. And many of them are gene therapy and I think most if not all of them are for rare conditions. And I think one of the questions asked about the single arm study, I mean actually, some of I guess the single arm study this person ask us is about the CAR T product and also the SMA, the XMA product, which I use the single arm, but I think for other products, we really think the concurrent control study is the more effective way to develop a treatment. And I think the Luxturnastory speaks on its own. Thank you.

Naomi Knoble:

Thank you, Lei. Lindsay. What gives you hope in rare disease product development?

Lindsey Murray:

Just the level of commitment I've seen from all of the stakeholders I engage with to get this done, to get, move, move the science forward. We're seeing it with C-Path, we're seeing unprecedented levels of data sharing and that really allows us to start looking at diseases. And rather than, it's hard to do (COA) development when you have 60 people in a trial, it's, you're not going to get a lot of good psychometric data there. But when you can pull data across multiple trials and natural history studies, we have examples in Friedrich's Ataxia, where we've been able to map out then how a tool that's being used in Friedrich's Ataxia differed based on onset of disease.

And that was only possible because we had seven studies that were pooled in access to 1300 patient level data sets. And that that's just so powerful. And really I get so excited thinking about what we can do with (COA) development, looking at how scales that are commonly used in pediatric trials, how Vineland performs across clusters of diseases that maybe have phenotypic or genotypic similarities. And I think there's a lot of excitement there for what we can do when there's more information available to us and it's not siloed.

Naomi Knoble:

Thank you. There are a lot of reasons for hope, Abigail. We are at time, but I want to hear from you what gives you hope in the rare disease space?

Yuqun "Abigail" Luo:

Yeah, just like my colleague, they said that we had an exponential increase in submissions, marketing applications for cell and gene therapy. And we had quite a few approvals over the last few years. And if you Google FDA biological approvals for a year, you can go to that webpage and see a lot of those. And I think now all these programs and FDA efforts, stakeholder efforts, I think that really gives me a lot of hope that it's a good time for patients for public health.

Naomi Knoble:

Thank you and David?

David Rousso:

Sure, I'll be quick and kind of echo what's already been said. Certainly there's just a lot more industry, there's a lot more biotech and pharmaceutical companies emerging in rare disease space, especially in the area of rare retinal diseases, inherited retinal diseases like the one that I was speaking about earlier. When Spark was founded in 2013, there were just a handful of companies in this space and now there

are quite a lot. And so this just increases the opportunities for the development of novel COAs that could potentially move beyond the more conventional measures like visual acuity where that has been challenging historically in the past for some of these types of rare IRD. So all in all, a lot of reason for help here.

Naomi Knoble:

I want to thank you all so much for a productive and engaging session. You've been wonderful and thank you for all that you're doing in the rare disease space. And with that, I would like to turn it over to my colleague, Dr. Laura Lee Johnson from the FDA for our fourth session today.

Laura Lee Johnson:

Hello everybody. And my name is Dr. Laura Lee Johnson. I'm the division director, a division director in the office of Biostatistics and FDA's Center for Drug Evaluation and Research. And I'd like to welcome you to our fourth and final session of the day, Considerations in Developing Rare Disease Endpoints: Multiple Endpoints with a Focus on Multi-component Endpoints.

In this session, we're going to be focused on discussing the use of multiple endpoints in rare disease drug development, and we're hoping that everyone will learn about the relevant multiple endpoint development challenges and opportunities with the focus on the multi-component endpoints. This session will end with a brief panel discussion. Dr. Knoble will be joining us back for that, where panelists will share thoughts on opportunities and challenges with such endpoint development.

So we're going to begin our session with a presentation. This time we're going to have three presentations from Kathleen Fritsch, a master mathematical statistician in the Center for Drug Evaluation and Research at FDA and a member of the multiple endpoints and clinical trials final guidance working group. Kathy, take it away.

Kathleen Fritsch:

Good afternoon. I'll be presenting on endpoint types and definitions. Next slide. So what is the definition of an endpoint? An endpoint is a precisely defined variable aligned with a research question that is analyzed statistically, usually comparing the test product to a control. And you typically need to specify the assessments of interest, the timing of the assessments, and what tools you'll use to make the assessments and in many cases how you plan to combine multiple assessments. Next slide.

Clinical trial assessments can usually be classified into four categories. One category is outcomes or events. And these are usually the adverse outcomes that we would like to delay or prevent through treatment like stroke or death. Another category includes signs and symptoms of a condition like pain or difficulty breathing, and many assessments fall into this category. And we've talked already today about some types of assessments that fall into the performance measure category, such as the distance walked during a time interval. And the first three categories are examples of clinical outcomes that can be used to provide evidence that a treatment provides clinical benefit in terms of how an individual feels, functions or survives. And the fourth category is biomarkers, which were also more fully discussed in an earlier session.

Key endpoint. Next slide. Key endpoints in clinical trials are usually classified into two main groups, the primary endpoints and the secondary endpoints. The primary endpoint family includes one or more endpoints, usually no more than two or three that are needed to establish the treatments efficacy and are aligned with the primary study objective. The results of the analyses on the primary endpoints determine whether the study met its objective, and then the secondary endpoint family includes additional outcomes that further characterize the treatments effects. The results of the primary and the

secondary endpoints are typically included in an approved products labeling and are used to communicate the key findings of a trial. Next slide.

Multiple assessments in a trial can be handled in a couple of different ways. One way is to define multiple distinct endpoints where each endpoint is based on a single assessment. Another approach is to combine the multiple assessments into a single score for an individual subject. There are many ways that assessments can be combined depending on the context. And the third approach is to define a composite endpoint. And the term composite endpoint is usually reserved for cases where the assessments of interest are the undesirable outcomes or events, and the interest is in whether a subject experiences any of these undesirable events during the trial. And next, I'll provide additional details on each of these three approaches. Next slide.

In some situations, it's appropriate to define individual endpoints for each key assessment. For example, if you have a disease with the key endpoints of pain and nausea, it may be appropriate to define one endpoint for pain and one for nausea and analyze them individually. The advantage of this approach is that the clinical interpretation is relatively straightforward and each endpoint is narrowly defined. However, for more complex conditions, this could lead to a long list of endpoints, which might require a larger trial to adequately assess them. One example where multiple distinct endpoints are used is in acne where there are three co-primary endpoints, and the three endpoints are based on inflammatory lesions, non-inflammatory lesions, and a global assessment. Next slide.

On the other hand, for certain conditions it may be appropriate to combine multiple assessments into a score for each subject and such endpoints are called multi-component endpoints. Multi-component endpoints may be appropriate for situations where the condition of interest has a variable or heterogeneous presentation across subjects and for other conditions, it may not be possible to characterize the condition with a single assessment, such as when an endpoint is based on activities of daily living is of interest. In some situations to ensure that the treatment effect is clinically meaningful, it may be important to demonstrate that individual subjects show improvement on multiple disease elements. Assessments can be combined in a wide variety of ways, but some examples include calculated sum scores or responder definitions where a subject is classified as a responder if they meet several different criteria.

One advantage with using multi-component endpoints is that you only need one or maybe a couple of endpoints to incorporate the key assessments. However, the complexity has a downside that it may be harder to interpret or identify which components are impacted by treatment or even if some of the components may be negatively impacted by treatment. Know for a multi-component endpoint, the conclusion regards the overall effect, and while it's usually important to look at the results for the individual components, you can't make conclusions about the individual components unless the trial is specifically and appropriately designed to evaluate the individual components and the components are meaningful and fit for purpose on their own. Next slide.

One example of a multi-component endpoint is a sum score such as the Montgomery-Asbury Depression Rating Scale (MARDS) for major depression score. The scale is based on 10 items that are scored from zero to six, leading to a total score that ranges from zero to 60 with higher scores indicating more severe depression. Another type of multi-component endpoint is the responder definition. For example, for fungal infections of the toenail or onychomycosis, a subject is classified as a responder with a complete cure if both the target toenail has no remaining signs of the disease, and if two types of lab tests are negative, indicating that the pathogen is no longer present. Next slide.

There has been confusion regarding the terms composite endpoint and multi-component endpoint. Historically, the term composite endpoint referred to a specific type of endpoint of which the most common example is the MACE endpoint, which evaluates major adverse cardiovascular events. For

example, a MACE endpoint could be defined as whether a subject experiences a heart attack, or stroke or death during the trial. This is an example where the objective of the study is to prevent or delay the occurrence of clinically important and related events. Thus, composite endpoints are structured differently than the examples I gave earlier for multi-component endpoints, which are used typically when the objective is to demonstrate improvement on symptomatic conditions.

This nomenclature distinction is useful because the cardiovascular community developed recommendations for analyzing MACE style endpoints and these recommendations were not always directly applicable to other multi-component endpoints, such as the recommendations that it's always important to look at and report the individual components. Using different terms for the different clinical context helps minimize confusion about the expectations for reporting. Next slide.

In summary, the endpoint selected should align with the study objective. The choice of whether to use multiple distinct endpoints or multi-component endpoint should be driven by the study objectives rather than just concerned with the overall sample size or analytical convenience. And the appropriateness of the choice to use the multi-component endpoint may depend on the complexity of the condition, the interrelatedness of the assessments of interest and the interpretability of the multi-component score. Next slide. And here are two sources of some of these definitions, the BEST resource and the multiple endpoints and clinical trials guidance. Thank you.

Laura Lee Johnson:

Thank you for that presentation Dr. Fritsch. In our next presentation, I'd like to welcome Lili Garrard, a master scientist and technical lead for the patient focused statistical scientist group in CDER at the FDA. Lily.

Lili Garrard:

Thank you Dr. Johnson for the introduction and good afternoon, everyone. I'm Dr. Lili Garrard and I will talk more about considerations with constructing multiple endpoints. Next.

So endpoint development itself is hard, especially in rare disease drug development. I think everyone can agree with that. It's challenging to assess a single concept of interest across all patients due to the heterogeneity within the disease. And it should also be acknowledged that there is no perfect endpoint strategy when a disease affects multiple aspects of feeling and functioning. It may be necessary to consider different aspects to adequately assess benefit, and we should consider the strength and limitations of various approaches. And whenever possible, we should evaluate several different endpoints in earlier studies to inform endpoint selection for later studies. Next.

So let's use an illustrative example to explain some ways we can think about constructing endpoints, building on what Dr. Fritsch just discussed. So suppose we have five different outcomes of interest that we think the medical product, the drug, will impact that are associated with a disease. Next. One approach is to construct separate endpoints for each aspect of health. In our illustrative example, we will have five separate endpoints corresponding to the five outcomes of interest. With this approach, sponsors should describe the role of each endpoint. Next.

So for example, we can have one primary endpoint and multiple secondary endpoints. This option might be helpful or useful. When there is one core or cardinal manifestation of a disease that most patients can be expected to experience and that is regarded by patients and/or caregivers as important, which can be assessed as a primary endpoint. Then the secondary endpoints can be created for aspects of health that might not be experienced by all patients and/or are viewed as relatively less critical, but still important to patients and/or caregivers. Next. Or we could construct co-primary endpoints. And this option may be appropriate when there are multiple aspects of health that are critically important to the

disease being studied, such that a treatment benefit can only be concluded if the medical product has an effect on each of the designated endpoints. Next.

So by creating a separate endpoint for each relevant aspect of health, there is clarity about which aspect of health has or has not been affected by the medical product because each endpoint corresponds to only one aspect of health. But there are several challenges with this approach that also should be considered. First for diseases with many possible manifestations, this approach may be challenging to use if it is not known ahead of time which aspects of health are most likely to improve as a result of using the medical product under study. And second, depending upon the roles of the endpoint, multiplicity adjustments might be needed necessitating a larger sample size to ensure sufficient statistical power. And finally, if patients differ from one another in their symptoms or functioning due to the disease, then the treatment effect estimated for any one endpoint will be diluted by the patients for whom the endpoint is not relevant. For example, patients who never had a given symptom cannot improve with treatment. Next.

So another very helpful approach discussed by Dr. Fritsch earlier is to construct a multi-component endpoint. Again, defined as a within-patient combination of two or more components, each reflecting a different aspect of health. Constructing a multi-component endpoint for an individual patient requires observation of all the specified components for that patient. Then a single overall rating or status on the endpoint can be determined according to a pre-specified algorithm. Next.

In order to construct a multi-component endpoint, we need to carefully consider the interpretation of the overall endpoint. To this point, the selection of individual components is absolutely critical. Some considerations include but not limited to; the clinical importance of each component, whether the different components would trend in the same direction within a subject, how each individual component will be measured, and we need to think about how will interpretation be impacted when combining different types of components. For example, if we were to combine a biomarker component and (COA)-based endpoint component together into a multi-component endpoint, how would we interpret the overall endpoint results? And then of course, the interpretation of the overall endpoint also depends a lot on the method on the scoring method used for the overall endpoint and each component including any weighting schemes if applicable.

Next, so on the slide, I list some take home messages from my talk just now. I want to reemphasize the point that there is no perfect endpoint strategy when a disease affects multiple aspects of feeling and functioning. So sponsors should really choose the best for their context of use and to help facilitate efficient communication and discussion between different stakeholders including FDA. A well justified rationale to support the proposed novel endpoint should be provided regardless of the choice of the endpoint. And this rationale should include, for example, the strength and limitations of the proposed endpoint and why the proposed endpoint is important to patients and or caregivers. And if a multi-component endpoint is being proposed, then sponsors should provide justification for the components included and the algorithm for combining them into the endpoint. And the last thing very important is to always keep the interpretation at the end in your mind. Thank you.

Laura Lee Johnson:

Thank you, Dr. Garrard. I want to welcome our final presenter, Kevin Weinfurt, a professor and vice chair of research in the Department of Population Health Sciences at Duke University Medical Center and a faculty member of the Duke Clinical Research Institute. Kevin is currently splitting his time between his work at Duke and working as a special government employee for FDA, supporting the drafting and implementation of the Patient Focused Drug Development Guidance series. Kevin, take it away.

Kevin Weinfurt:

Thank you very much, Dr. Johnson. I'm going to say just a little bit more about multi-component endpoints. If we could go to the next slide. As Dr. Garrard was saying, a multi-component endpoint is this within subject combination of two or more components. First, I just want to say a few things about how these different components might relate to measures in the study.

If we go to the next slide, one option here is that each component could be the score from a different (COA), right? In this case, we've got the Symptom A severity measure, Symptom B, severity measure, and so on. And those scores from different measures might be combined in some way to construct an endpoint. We go to the next slide. A second option here is that each component could be the score from a subdomain of a single multidimensional clinical outcome assessment. Some measures are designed to generate a profile of scores. So in this case, this hypothetical measure, the ABC functioning measure has five subdomains. A sponsor might choose to combine those if there's good justification for doing so into some overall value that is the multi-component endpoint. And the third option, if we go to the next slide, oh, I'm sorry, we got one more here. Each of these subdomains themselves could be assessed with multiple items or tasks, right?

Okay. Now we're ready for the next slide. It's option three here, where each component could be the response to an item or task from a single measure. So in this case, the measure is designed to look at the disease X symptom severity, and it has five items here that are each asking about five different symptoms. If we go to the next slide here, this type of measured design is referred to in the PFDD Draft Guidance Three as a composite indicator model. That means that the separate items or tasks are all grouped together like ingredients in a cake to sort of define the thing that you'd like to measure, as opposed to a reflective indicator model where each of the items are supposed to be reflections of or caused by the same underlying thing. So sometimes you've got these composite indicator models underlying a measure and you could choose then to analyze that overall score and see that as a multi-component endpoint because each of those five items are really components that are being combined to arrive at some overall value.

So we go to the next slide there. So as Dr. Fritsch and Dr. Garrard noted, there are some benefits here. A multi-component endpoint has the potential to evaluate the entire range of important disease manifestations. No multiplicity adjustment is needed and it can be very efficient if the treatment effects on the different components are generally concordant. The challenge here though, is in trying to justify the weighting, or the algorithm that's used to combine the components. And that's a very important part of the rationale for the multi-component endpoint is justifying that. And I wanted to just touch on one approach that might be considered and we want to use an example from a 2011 paper if we could go to the next slide.

This is actually more of a MACE endpoint, and I'm using this example to illustrate a process that was used. It was in the ASSENT-3 trial and, if we can go to the next slide, they had these different components of death shock, congestive heart failure, and repeated MIs. And they wanted to figure out how to weight these to combine them in such a way that that MACE endpoint was more clinically meaningful. And so, they used a survey procedure to elicit weights on a 0 to 15 point scale for individual endpoint components from 23 experts. These are cardiovascular experts, but you could use this type of approach with any type of stakeholder, including patients, find out what's relatively important to patients.

And their primary analysis used the multi-component endpoint here that was constructed using the median weights from these survey respondents. But now, here's the important point that I wanted to stress, they also conducted sensitivity analyses. And so, when they did this survey procedure, you can

see from the figure that they got a range of weights from the different stakeholders who participated. So there was this range. They used the medians to define the endpoint for the primary analysis, but you'd like to be assured that the trial findings would be robust against any reasonable variations in how those weights are done.

And so, they did a Monte Carlo simulation of different versions of that multi-component endpoint where they varied the weights within the range of values supplied by the survey participants. So they didn't just vary them between 0 and 15 points. They varied them, for example, with shock between looks like about 3 and just over 6. So they tried to reflect the range that they heard from their experts and they empirically derived confidence intervals then around the treatment effect based on 1,000 repetitions in this Monte Carlo simulation.

But this specific example just illustrates that, first of all, it's good to go to the stakeholders, primarily the patients, to try to understand the relative weighting when possible. And that can be done in a systematic way. And that we can do some explorations to see how robust the findings are, if there's some variation in the weighting that's being used. And there we go. That's all I wanted to share today. So Dr. Johnson, I'll hand it back to you.

Laura Lee Johnson:

All right, thank you Kevin.

So we'll now move on to our moderated panel discussion, and I want to welcome back Dr. Naomi Knoble who moderated Session 3. Naomi, do you want to add anything before we move on to Q&A?

Naomi Knoble:

Yes. Thank you so much Laura Lee. And first, I just want to thank my colleagues, Doctors Fritsch, Garrard, Weinfurt, just for the thoughtful presentations here. That was just fantastic.

Yeah, I want to highlight two points here. One, that endpoint development is a paradigm shift from what we do clinically. It's a strategic approach, it's a balance of many factors that Dr. Fritsch touched on this in terms of the context of use. You have to balance the many factors that comprise that umbrella term that we often use here. And clinical trial endpoint construction requires that different approach from assessments with patients in clinical practice to measuring outcomes successfully in clinical trials and constructing endpoints. And as I alluded to earlier, the FDA has a number of guidances around this.

The second key point I want to make is in regards to the totality of evidence, especially in the rare disease space. Dr. David Rousso touched on this in his session presentation that was just prior to this one. And he commented on how it was having relevant strong secondary endpoints and exploratory evidence as well, plus the primary efficacy endpoint that the measurement itself though had some limitations. And that wasn't a deal breaker because of the totality of evidence. And as my colleagues here have indicated, there are many ways to construct endpoints, but especially in the rare disease space, all the evidence that's brought to bear on interpreting the clinical benefit of a product in the context of use is what's most important here, I think, is we're looking at how to interpret products in the data in front of us. So thanks Laura Lee.

Laura Lee Johnson:

And thank you Naomi, and everybody else.

So I want to welcome everybody. So we have Naomi and our three presenters, Kathy Fritsch, Lili Garrard, and Kevin Weinfurt. So if we can go to the next slide. Thank you.

I want to briefly turn us to the list of discussion questions that we have for the session. And you'll also find those questions are going to be posted here in a moment in the Zoom chat. And we can start with the first one. And I'm going to start with you, Lili. "What are some of the biggest challenges that stakeholders may experience in developing and implementing multiple endpoints, and in particular multiple component endpoints for rare disease research?" And can you comment on any strategies that might be effective for overcoming or minimizing the impact of those challenges? That's kind of a big overarching question, so if you just want to take one little part of it, that's great.

Lili Garrard:

Yeah, sure. Yeah, that is a big question. There are certainly a lot of challenges. So I may just focus on... During my work, I focus a lot on working with clinical outcome assessments, COAs. So, in terms of rare disease development, the measurement challenges is a huge thing for us. And I myself as a statistician. So I often say that if I have a simplified wish is that I would have these high quality measures so that I could generate, or valid, reliable data can be generated from those measurements. And then I can turn it into some kind of endpoint that could be interpreted in the end. But we often get stuck at the first step, which is to have high quality measurement.

So I think some of the things that sponsors really should consider is, first of all, it's been touched a lot by previous presenters already. Make sure you talk to the patients. We understand, we completely sympathize the situation that we do not have a lot of patients to work with in rare disease space. So make sure you utilize all the opportunities you have to talk to them. Understand what they would like to see in trials. And then, just try to strike a balance between what the patients consider to be important and meaningful versus what your medical product can impact.

And then, the other thing is make sure that you have standardized ways to collect the data. We often see a lot of challenges in this area. For example, in a utilization of clinician reported outcome, many, many times we have multiple raters. So here, as a statistician, I'm going to ask the question, do we have evidence for inter and intra rated reliability? I want to make sure that when a person a, selects a score of two for one patient, it's the same meaning as a score of two rated by a different clinician because all of that will contribute to the overall interpretation of the results. So I may just stop there and let others chime in. Thank you.

Laura Lee Johnson:

Do others have comments or thoughts?

Naomi Knoble:

Sure. I'll piggyback onto Lili's comments here. I think Lili, your point about standardization is well taken and we, certainly, do see challenges on the implementation side of clinical outcome assessments, and also too, just to be fair, other forms of assessments too. And sometimes it might even just reflect a sponsor's internal skillset. Sometimes the clinical operations aspect of running clinical trials is quite a separate team than the clinical team who's helping to design the protocol and selecting outcomes. And so, sometimes there can be that schism, but certainly there's fairly robust research that addresses implementation challenges that we commonly see. And that's, certainly, something that we need to attend to more as we look at especially these more complex forms of endpoints, and not just losing components of it due to lack of standardized implementation. Thank you.

Laura Lee Johnson:

Any other thoughts?

Okay.

Have another question. And Lili, I'm going to start with you again. What are some of the general tips, challenges, interpretation goals when you're developing, or using a multi-domain responder index?

Lili Garrard:

So multi-domain responder index, I think I saw a similar question in the chat also mentioning that. It's also referred to as MDRI, as a lot of us have heard about it. On this point, from my own perspective, I think in principle MDRI is not a bad idea at all. It's a way to include different assessments, and that's important to patients, and that's impacting their lives. But the challenges that I often see on my end is that I haven't seen one that has been successfully operationalized, and that is due to many different factors.

One is the components selected to put into the MDRI, I mentioned earlier in my presentation, how do you think about the interpretation when you combine a biomarker and a COA? So, in practice, biomarkers are going to change much faster than what you can observe from symptoms or functioning. So for example, we see biomarkers combined with, for example, something measuring cognition in a trial that's six months. Do we reasonably believe that we will see a change in cognition? So the endpoint results, the overall MDRI endpoint results are going to be mostly driven by the change in the biomarkers. So at that point, do you still consider this as representing how patients feel or function?

And then so MDRI takes on many forms. There are many variations. So one popular one is defining a responder for each individual endpoint. And then together use some kind of algorithm. For example, if any one component is a responder, then the overall endpoint is a responder. So that with my example of combining biomarker and COA together, if the change is going to be driven by the biomarker.

And the other one is if we say it is appropriate to define responders, it's very challenging in practice to identify the cutoff, or the thresholds, the responder thresholds, because it needs to be pre-specified. Those thresholds needs to be pre-specified And then with justifications provided. In rare disease, it's extremely hard to have a lot of prior data, or from natural history data to justify those thresholds. And then, that together with the small sample size, et cetera, everything, it just makes it very hard to interpret in the end.

I don't know if others want to chime in. I clearly have a lot of thoughts about this. So I'm still waiting for the one example where MDRI truly works and I'll be rejoicing.

Laura Lee Johnson:

So it sounds like a lot of opportunity, but a lot of challenges and things that have to be worked out, which I think it sounds like can be an issue with any of the multiple endpoints, perhaps.

I don't know if others want to comment. Or if you also want to segue on some of these questions that also come up. I know one thing you brought up Lili is that intersection that many have biomarkers and COAs as well. So Naomi, did you want to comment on any of this?

Naomi Knoble:

Sure thing. Yeah, just to tag along to Lili's comments here and then pivot to this question of those challenges of incorporating so many different components into a single endpoint, it's a matter, it seems, of ensuring in these heterogeneous rare diseases especially that there is a substantive portion of the patient population. Certainly, we're not going to get 100% of patients necessarily showing improvement or change on all components, but that there would be a sufficient prevalence, I suppose for lack of a

better word, of symptoms, or impacts, or whatever the thing is that you're measuring that would all patients could at least contribute to the many portions of a multi-component endpoint.

And as Lili pointed out, and I think as Kevin pointed out in his presentation as well, it's just unfortunate that not all patients will necessarily evidence improvement on all aspects of these multi-component endpoints. And that just leads to missed measurement opportunities, unfortunately.

And to speak to some of the challenges of integrating biomarkers and clinical outcome assessments into the single multi-component endpoint, I think one of the biggest challenges that we see is that there may not necessarily be enough evidence to suggest that all pieces should be included in a multi-component endpoint. We all do know that not all, maybe biomarkers and COAs that the evidence doesn't always track in the same direction, but if there is enough of a justification and rationale for including both together, sometimes that rationale can be the evidence needed.

Kevin, were there other thoughts that you had on that?

Kevin Weinfurt:

Sorry. No, I didn't have anything to add to that one.

Laura Lee Johnson:

So I am going to use this though as a pivot to another one of the questions that we have in Q&A that was saying, "Without external validation, how can you really justify the weight assigned?" And so is that one, Kevin, that maybe you want to start us off with?

Kevin Weinfurt:

Sure. It's a great question. And in that example that I presented that 2011 paper from the ASSENT trial, I didn't share that they actually did a second survey of experts as a validation to confirm the medians and ranges of values they had there. And so, when you could do that type of thing, I think it would be really useful. But I think we also have to recognize that there may be systematic processes for developing weights that are drawn out a very small patient population, especially in rare diseases. And it may be a process like a Delphi process where you've got some systematic consensus building process about what the relative weights ought to be. And you could still get some sense of the variability from that process as well. But I think the process of deriving weights that reflect people's relative values isn't quite like a parameter estimation exercise. It's just trying to get some systematic way to reflect the relative views that people have and conduct sensitivity analysis with that. So might not always be possible to get that external validation, but when you could, I agree, that would be very valuable.

Laura Lee Johnson:

Other thoughts?

Okay so, I want to make sure we have time to discuss I think the fifth question that was listed in the chat. Kathy, can you talk to us a little bit about the interplay between the measure, the assessments, the endpoint, the analysis, the interpretation? I think that's a common theme that we've heard today. So can you start us off on that?

Kathleen Fritsch:

Sure.

I think the individual assessments are the sort of building point. That's what you're actually recording in the chart or on the clinical forms that you're filling out for the trials. And then, you have to think about,

well, how do I want to combine these into a single endpoint? And we looked at examples like some were sumscores. So that's one way of approaching it is we talked about we'll look at different ways of weighting it if we don't want to do the simple straightforward weighting.

So then you're thinking about, well so I have this score now, what is the best way to approach it? How will it be interpreted? We'll be looking at means, or do we need to define some sort of a threshold that will make a responder definition? All this sort of goes into play with what the objective is, what we know about the scales, can we interpret changes of a certain size, or just changes in general? And that will lead us to then, the types of analyses that are appropriate for however we end up constructing the endpoint.

So there's so many ways to put these together that it's helpful to keep in mind, what are the impacts of the types of measurements? How can we interpret them when they're put all together? Do we know what it means to have a five point change on this particular scale or whatever? And how can we present that to patients and healthcare providers so that they can understand what the impacts are? And that all goes into also the total, how easy it is to get an endpoint based on the number of subjects available, which is particularly challenging for rare diseases.

Laura Lee Johnson:

Do others want to comment?

Naomi Knoble:

Yeah, Kathy, I agree and want to just echo all of your points, but without repeating them.

I would just add to it that, again, in the rare disease space the totality of evidence is so important. And outside of the endpoint strategy, there are other supportive forms of evidence that can be brought to bear on especially that point around interpreting evidence and is it clinically meaningful? And one way to do that are through exit interviews, or what people are also calling in-trial interviews with a baseline and end of treatment period prior to un-blinding, or unmasking patients. And those interviews with patients can really help bring patient perspectives the lived experience of a product, its benefits and risks for patients and families to directly comment on what they think the experience has been like, and if they think the change is meaningful. Or, as we often I think see in the rare disease space, if stabilization of symptoms was meaningful, which is an interesting way to look at scores and such.

Laura Lee Johnson:

And I do know that... Oh, sorry, Naomi, did you-

Naomi Knoble:

You go right ahead. No, Laura Lee, go right ahead.

Laura Lee Johnson:

I know that many of our statisticians have been very involved in a lot of our patient listening sessions, and really taking a lot of that information to heart to think about, not just about design and endpoints, but also thinking about how should they be planning those analyses.

So I know we had one more question, but I think we're actually at time. So did anybody have a burning comment that they wanted to make?

Okay. Thank you all.

And, at this point, I want to thank everybody for our productive and engaging session. And I know that while people were short, that you all could talk for days about this. And luckily, we have a few years with our data to actually have these continued discussions. And now, I'm going to turn this over to Dr. Mark McClellan, who'll share some closing remarks prior to adjourning our meeting for today.

Mark McClellan:

Thanks, Laura Lee, and I want to thank all of you for some great discussions today, a bunch of stimulating ideas, and potential path forward on these important topics related to rare disease endpoints.

Just a few of the many ideas that were discussed included things like ideas for partnering and collaborating with patients early on. That's a central theme that we heard throughout the digital technology session, and it's come up in other parts of the meeting as well. Very important that the use of digital health technologies make patients' trial experience easier and more meaningful. Also, ideas on validating new measurements and assessments is a process that's going to take time. As Laura Lee was just saying, it will require a lot of contributors and opportunities to engage with physicians, programmers, statisticians, regulators, other relevant specialists will help make this process go forward effectively, and as efficiently as possible.

With regard to development of biomarkers as surrogate endpoints, we heard that because of the nature of rare disease populations with small population sizes, data is often limited. So figuring out the appropriate means for sharing data, and make the most of the limited patient experiences is very important. Also, there's no single recipe to ascertain biomarker clinical evaluation. Expanding communication with FDA whenever possible can help move these efforts forward.

With clinical outcome assessments, we heard that it's important to have at least some understanding of the natural history of a disease going in, a depth of... or qualitative study data or the depth of the qualitative study data also helps even if you can't clearly quantify the impact over time through a well-known detailed natural history.

Finally, with multiple endpoints, instructing endpoint development is difficult. There's no perfect strategy when diseases affect multiple functions for heterogeneous diseases as is often the case with rare diseases. Combining multiple primary and secondary endpoints into a single multi-component endpoint may be a useful approach but, again, care to make sure that if these are really distinct endpoints that we're paying attention to those too, care in this process is critical.

Overall, we heard throughout the sessions today about the importance of focusing on areas of unmet medical need for patients and establishing robust collaborations between different types of stakeholders.

Tomorrow we're going to continue our discussions starting at 1:00 PM Eastern Time. We've got a number of sessions teed up with speakers and an opportunity for exchanging ideas about the RDEA pilot program then. In the meantime, I'd like to thank once again all of you who contributed today. Also, one more request that anyone participating in this meeting can submit written comments regarding the event to [regulations.gov](https://www.regulations.gov) until July 23rd, that's [regulations.gov](https://www.regulations.gov).

So for now, have a great rest of your day and we look forward to seeing all of you tomorrow. Thank you very much.