Primary Mitochondrial Diseases:
A Rare Disease Virtual Workshop
May 22, 2025 | 10am-3:30pm *(eastern)*

Morning Transcript

Welcome
Susan C. Winckler, RPh, Esq., Reagan-Udall Foundation for the FDA

Susan Winckler: Hello and welcome. Thank you for joining us for this public meeting. I am Susan Winckler and I serve as Chief Executive Officer for the Reagan-Udall Foundation for the FDA. For those of you who may be new to the Foundation's work, we are the nonprofit non-government organization created by Congress to help the FDA do more to promote and protect the public's health. One way we do that is by convening meetings such as this one to help [00:00:30] the agency share information and to hear from stakeholders about important issues. These engagement opportunities help inform the agency's work.

Today, we are pleased to be collaborating with the FDA to host its workshop on primary mitochondrial diseases. Of note, the foundation does not advise the FDA on nor in any way engage in regulatory decision-making. Today's discussion is focused on education and information sharing among the community. We [00:01:00] will not address FDA-specific regulatory action.

Before we begin, I need to cover a few housekeeping issues. We have several hundred virtual participants joining us today, and thank you so much for investing your time with us. But because of the size of the meeting, attendee cameras and microphones will remain off throughout the event. We do, however, want you to engage with us. So to ask questions or share your comments, please use the zoom Q&A function and I will do my best to ask [00:01:30] as many of your questions as we can. We are recording the full meeting and we will post the video recording along with the slide deck and a transcript on the foundation website, which is Reaganudall.org, next week.

Before we dive into the agenda for the day, I thank our colleagues at the FDA Center for Drug Evaluation and Research. We appreciate your partnership and support in planning this vital public meeting. So let's think about our time together today. Throughout the day, academic researchers, [00:02:00] patient advocates, clinicians and industry representatives will discuss the latest developments in our understanding of primary mitochondrial diseases and explore the future of therapeutic development. We will begin with an overview of primary mitochondrial diseases, focusing on the clinical presentation,

underlying pathology, and the current challenges associated with diagnosis and treatment.

We'll explore key aspects of drug development, including the selection of patient-focused outcomes and relevant statistical [00:02:30] considerations. After a mid-meeting break, we'll hear about current approaches, challenges, and opportunities in therapeutic development. Before inviting our closing panel to share their ideas about moving forward, they're going to help us reflect on and think through all that we have heard throughout the day. So that is the housekeeping and helping us understand what is ahead of us for the next few hours.

I want to now turn the stage over to Dr. Catherine Pilgrim-Grayson. [00:03:00] She serves as the director of the Division of Rare Diseases and Medical Genetics within the Center of Drug Evaluation and Research at the US FDA, where she provides scientific, clinical, and technical direction on all medical and scientific decisions and judgment in connection with the review and evaluation of drugs in that arena. Dr. Pilgrim-Grayson, welcome. And the floor is yours.

FDA Opening Remarks

Catherine Pilgrim-Grayson, MD, MPH, Center for Drug Evaluation and Research, FDA

Dr. Catherine Pilgrim-Grayson: Thank you so much, Susan. Good morning and welcome everyone. We are thrilled [00:03:30] for this opportunity to engage with all of you today. And as Susan said, we in the division of Rare Diseases and Medical Genetics, the Rare Diseases team, along with the Reagan-Udall Foundation are pleased to convene this workshop to explore opportunities to optimize therapeutic development addressing primary mitochondrial diseases. We aim to discuss best practices and share lessons learned in designing and interpreting clinical studies that could accelerate drug development. FDA is committed [00:04:00] to doing all we can to further the development of drugs for rare diseases, including primary mitochondrial diseases. We want to alleviate the profound impact on patients and families.

We share the goal of bringing effective and safe treatments to patients as quickly and as efficiently as possible. We do recognize the challenges facing patients and families and also the challenges facing drug companies pursuing development of treatments for primary mitochondrial [00:04:30] diseases, or PMD. This workshop aims to address these issues and turn lessons into opportunities for continued work together to expedite drug development in this space. This day calls for a big note of thanks. Thanks to the Cedar Arc Program for funding the workshop. Thanks to all who've worked so diligently in planning. This event has been over a year in the making.

I would especially like to commend Lea Ann McNee, Esther Brodie, [00:05:00] Perpetue Backer, Hannah Gray, and Susan Winckler at the Reagan-Udall Foundation. And I want to commend my colleagues at the FDA. You'll hear some

of them speak today, so you'll get to know them a little bit. But I would also like to recognize those who labored behind the scenes, including Yuliya Yasinskaya, Maura Ruzhnikov, John Idso, and Meshaun Payne. Special thanks to Dr. Amal Khaira and Dr. Philip Jeske, who spurred us on to have this session. I'm grateful for our presenters and [00:05:30] panelists and I'm drawing special attention to the patients and caregivers who are participating today.

You are the drivers of our work and your input at every step of drug development is of paramount importance. Thank you for all that you do to advance this work. We're looking forward to the discussion and we're building on previous experience. So as many of you may recall, FDA hosted a Mitochondrial Diseases Symposium in 2019. We covered topics such as [00:06:00] integrating mitochondrial biology in the design of drug development programs. We discussed lessons from previous programs, sharing about what worked, like the establishment of registries, exploration and refinement of patient focused outcomes, and what did not work. For example, putting a laser focus on the six-minute walk test as a single outcome measure.

We also talked about how best to address regulatory and scientific considerations when designing trials. Since that symposium, we've continued to work on developing [00:06:30] and refining endpoints and thinking deeply about child design, implementing additional learnings from completed trials. So today is a perfect opportunity to share our accumulated knowledge and use it as a springboard. As we all know, this has not been an easy road. Challenges we face in designing clinical trials for PMD include the fact that disease definitions are ever evolving. Numerous genes cause the diseases. There are small numbers of patients for each genotype, [00:07:00] and patients have heterogeneous presentations. The natural history is not always clear.

Further, pediatric patients with rapidly progressive disease often have early mortality, adding to the complexity and child design considerations. Moreover, no specific and sensitive outcome measure can be used across all PMD. It's clear then that a one size fits all approach does not work in this group of diseases. And you'll hear people say that repeatedly today. We need to [00:07:30] be nimble. Collaboration between researchers, clinicians, and patient groups is vital to designing clinical studies that address the unmet needs of individuals with PMD.

This is why we're convening this workshop: to build on our collaborations. By engaging with all stakeholders, we aim to leverage the patient experience and perspective, leverage expertise and innovation, strengthen dialogue and increase understanding of the regulatory frameworks so that we can tailor approaches to each disease. [00:08:00] As Susan said, today we're going to hear from academicians, clinical investigators, patients and their representatives, drug companies and FDA on patient population considerations, development and selection of patient-focused outcomes, statistical issues, and current approaches and opportunities in drug development for PMD.

We'll round out the day by taking what we've learned to continue moving forward. So with that, it's time to turn to session one, setting the stage. So Susan, I'll turn it back to you to [00:08:30] get us started. Thank you.

Clinical Trial Design and Implementation: Patient Population Considerations

Anna Choe, MD, MPH, Center for Drug Evaluation and Research

Susan Winckler: Excellent. Thanks so much, Dr. Pilgrim-Grayson. So I will note, as happens sometimes in these meetings, our first speaker has not yet joined and so we are going to turn and first hear about a very important component of our discussions and with that, and that is clinical trial design and particularly as we think about the trial population and considerations for that. [00:09:00] So to help us in that introduction and that grounding, Dr. Anna Choe, who is a clinical team leader within FDA Center for Drug Evaluation and Research is going to help us think through those components for design and clinical trials. So Dr. Choe, we are excited and interested in what it is that you have to share. I'm going to step away so that we can see your slides and you, and hear your great content.

Dr. Anna Choe: Thank you, [00:09:30] Susan. Let me make sure that I can control the slides. Good morning everyone. My name is Anna Choe. I'm one of the medical officers in the division of rare Diseases and medical Genetics in the Center for Drug at the US FDA. And today I'll just be discussing population considerations as it pertains to clinical trial design in primary mitochondrial diseases. So to start, I'd like to note that this presentation [00:10:00] represents my own views and does not represent an official FDA position. I have no financial interest to disclose and in my talk, the term "drug" reflects both drugs and biologics.

This is a roadmap for my talk. I will start by setting the stage, going through some of the regulatory framework terminologies so that we have common terminologies that we use in the presentation. I'll go through some of the considerations and population selection, then share recent approvals [00:10:30] in our division and highlight two case examples as it pertains to population selection. So to be approved for marketing when an application comes in, the review team has to determine that the drug is safe and effective. And effective is codified in statute as demonstrating substantial evidence of effectiveness for the proposed indication. On the other hand, safety is not explicitly defined in statute of regulations, but because all drugs can have risks, [00:11:00] this is really interpreted in the setting of benefit risk assessment where the benefit outweighs its risks.

So then what is substantial evidence effectiveness? And bear with me, some of these terminologies are sort of dry, but it's really important because I'll use these terms again later in the presentation. So effectiveness established with substantial evidence. And in the 1962 amendment, this was interpreted as having a minimum of two adequate and well-controlled studies or [00:11:30] a single study that is essentially equivalent to having two adequate and well-controlled studies, a multi-center statistically persuasive study. And the reason,

the scientific principle underlined the requirement for two is because a single study can have systematic bias. And having a second study replicating these findings can increase our confidence in its conclusions. Subsequently, there was complementary statutory standards established in '97 that clarified that an adequate [00:12:00] and well-controlled study, one of them with confirmatory evidence can constitute substantial evidence.

So then what's an adequate and well control study, what we call AWC? AWC studies are studies that are designed well enough so that you can distinguish a drug effect from other influences that can happen spontaneously or due to a placebo effect or bias observation. And in general, the effects seen from an AWC must be considered to be clinically meaningful to [00:12:30] constitute substantial evidence. And these are the different features of an AWC. You can read through the list, but in general, clear statement of objectives, a design that allows a valid comparison with the control. Inclusion criteria, exclusion criteria that assures that the subjects in the trial have the condition that's being studied and measures that minimize bias, having well-defined reliable methods to assess a response and adequate analysis plan.

[00:13:00] So as you all know in general, RCTs randomized control trials are considered the most rigorous study designed as an AWC. And this is because of the randomization and blinding that assures that we minimize bias in the study. Typically, we think about parallel placebo controlled studies as an RCT, but of course there's randomized withdrawal or crossover studies that can also be considered an RCT. Alternatively, you can also consider [00:13:30] baseline controlled or externally controlled studies to demonstrate substantial evidence of effectiveness. But these studies typically require a well-defined natural history, predictable natural history, an endpoint that's not prone to bias and a robust treatment effect. And there's been a growing interest in this area as well.

So PMDs are rare conditions. [00:14:00] It has a complex genotype, can present as a multisystematic disease, have heterogeneous varial presentation. Some of the available natural history have been shown a lot of predictable natural history course and poor genotype phenotype correlation that limits some of the study designs that I mentioned earlier. So for PMD, a careful selection of endpoint and trial population is key to optimize the success of an AWC. And this isn't something new. The PMD [00:14:30] community has been aware of this. Sponsors, investigators and patients are aware, and this is something we've all been thinking about working on, but as [inaudible 00:14:40] McCarthy will share in the next presentation, there also have been some challenges.

So for endpoint selection, there are several considerations go in and we spent a lot of time discussing with sponsors, trying to figure out which end point would be adequate for the trial. So we [00:15:00] think about the drug product, the mechanism of action, the non-clinical efficacy data, the exploratory trial. If the natural history is available, we also think about whether it's sensitive enough to change within the trial duration and whether the end point would be predictable based on a natural history. The course of the end point would be

predictable. Of course, patient experience is extremely important and the PMD community has done a fantastic job with this, engaging with the patient community, [00:15:30] identifying symptoms that are relevant to them and determining meaningfulness.

And then measurement tool, the ability of reliability is another consideration. And then last but not least, effect size is also something that we consider as we think about the power. So similar considerations go into population selection, the drug products mechanism of action or what is known about it can inform that the population selection perhaps [00:16:00] be based on the molecular alterations or look at targeted symptoms. It can also inform whether we can enroll pediatric patients as more data is typically required to justify the risks for this vulnerable population. The natural history can help define the population as well. If you have a disease subtype that has a more predictable natural history or if we understand if the disease is going to worsen or if the goal is to stabilize, you may want to focus [00:16:30] on more milder population or more severe population with your inclusion criteria and enrichment.

Patient experience also ties into that where you want to think about the ultimate functional impact and how we will do the meaningfulness interpretation at the end. And then the measurement tool sometimes are validated in adults or specific age ranges in children, so that also can inform your population as well. And last but not least, sample size is a big consideration [00:17:00] when it comes to population selection, as I'll discuss in the next slide. I want to caveat by saying this is not meant to be comprehensive. There are some safety considerations that go into population selection that I did not include here. So in the 2019 presentation at the workshop for PMD, Dr. Karaa summarized this challenge really nicely between balancing between heterogeneity and the sample size. So if we reduce the variability to optimize the treatment effect and detect [00:17:30] it, then we were concerned about reducing the sample size and trying to achieve that balance is something that a lot of us in the PMD drug development space has been trying to figure out how to balance well.

So going onto some real examples, I included six recent approvals by our division. I excluded one that did not have a publicly available review published yet. Two products were approved in Neimann Pick C, one in Late-onset Pompe disease, one in Fabry disease, one in Alpha- [00:18:00] mannosidosis, and one in Acid sphingomyelinase deficiency. And as you can see from the quotes that I copied from the reviews, all these diseases are considered to be heterogeneous with variable presentations. Something else that I noticed when I looked through these reviews were all of the substantial evidence packages for these six products included at least one randomized controlled trials.

The sample sizes were small, [00:18:30] the largest one was 123 subjects, and the rest were mostly less than a hundred with the smallest sample size being 25 subjects. I do want to note that these diseases are not the same as PMD. They may be less heterogeneous and some of the products may have a more

targeted mechanism of action or have more available data to inform the population and the endpoint. So one size fits all does not work in rare diseases. But what I do want to note [00:19:00] is a strong study design such as an RCT can help overcome some of the uncertainties that we note when we have a trial in smaller heterogeneous populations.

So going onto two specific examples that pertain to population selection, Olipudase alfa was approved based on a single adequate and well-controlled trial with CE. They conducted a 52-week randomized blind to placebo controlled trial [00:19:30] in adults and assessed lung function and spleen liver size. The trial population for this was adults, ASMD type B, and they did enrichment based on the primary endpoints. And they focused on the ASMD type B versus A based on the differing natural history. Ultimately, the indication was approved for treatment of non-CNS manifestations of ASMD in pediatric and adult patients. And I wanted to note that the indication includes ASMD [00:20:00] type A in pediatric patients given the mechanism of action, disease pathophysiology and the available clinical data to support this. And this is discussed in the publicly available review.

The second example is Levacetylleucine, which is also approved based on one AWC with CE, confirmatory evidence. They also conducted a RCT. This was a crossover study in 60 subjects, and they assessed the functional Scale for Assessment and Rating of Ataxia (fSARA), [00:20:30] which was an endpoint to assess Ataxia specifically. The trial population here included pediatric and adult patients with cerebellar ataxia from NPC. And again, they enriched based on the primary endpoint. And they selected cerebellar ataxia specifically out of all the manifestations of NPC based on the known drug effects and their exploratory trial.

The indication ended up being treatment [00:21:00] of neurological manifestations of NPC in adults and pediatric patients. And again, I wanted to note that the indication, it goes beyond ataxia. It includes neurological manifestations, and the rationale given is because of different aspects of pragmatic neurological function that was measured by the fSARA. So before I end the talk, I wanted to share a few guidance documents that are relevant to population selection. And I wanted to highlight three in particular.

First one is Developing Targeted [00:21:30] Therapy in Low-Frequency Molecular Subsets of a Disease. This guidance discusses different source of data that you can use to group patients. It also talks about conducting your primary analysis in a specific population, but enrolling a broader population to do a preliminary efficacy assessment. The Enrichment Strategies Guidance talks about different enrichment strategies, but more broadly, it also talks about ways to reduce variability such as lead-in periods or assessor trainings. [00:22:00] And then the last guidance document I wanted to mention is the Conducting Clinical Trials with Decentralized Elements, which would allow improved enrollment and retention, ensuring your larger sample size. However,

the guidance also talks about introducing increased variability, so there are some pros and cons to be considered.

And I also wanted to highlight in the next session for the endpoint session, Dr. Naomi Noble and Dr. Yen Wang will also talk about endpoint topics that are relevant [00:22:30] to enrolling a heterogeneous population and reducing variability. So in summary, recent approvals in rare heterogeneous diseases were based on smaller RCTs. Natural history, mechanism of action, preclinical data, exploratory trial, patient experience, measurement tool, sample size, can all inform population selection. And the PMD community has done tremendous work engaging with the patient community, [00:23:00] publishing the voice of the patient reports, working with the academic community, looking at the non-clinical data and developing natural history studies. Nevertheless, PMD continues to have unique challenges and opportunities for selecting endpoints and trial population and trial design.

And one size fits all in PMD does not work. What has worked in another PMD program may not work in another one and what hasn't worked in another program may work in a different one. So it really needs to be tailored to the specific drug development [00:23:30] program. Lastly, additional strategies can be considered to reduce variability and improve enrollment and retention as well. So I really want to thank the organizers for this workshop today and I appreciate your attention so far and I'm looking forward to learning from the rest of the day. Thank you.

Setting the Stage: A Variety of Perspectives

Michio Hirano, MD, Columbia University Medical Center

Susan Winckler: Excellent. Dr. Choe, thank you so much. I'm reminded every time that I hear a, as always, [00:24:00] brilliant presentation from our colleagues at FDA of the complexity of the work that you discharge and then appreciate so much the ability to think through the various animating factors there, as well as the availability of the guidance documents, which helps us navigate that. So thank you so much, Dr. Choe. We will look forward to seeing you after the rest of our presentations [00:24:30] when we come back for a panel discussion. So thank you.

And now let's turn to Dr. Michio Hirano, who is the Lucy G. Moses Professor of Neurology, chief of the division of Neuromuscular Medicine and director of the H. Houston Merritt Neuromuscular Research Center at Columbia University Medical Center. Dr. Hirano, we are looking forward to hearing from you as you lay the foundation for the meeting by presenting the clinical presentations, [00:25:00] underlying pathology and more about this collection of diseases. And I know if there's one thing that I have learned from the rare disease community at large, it's that we are flexible and adaptable. So I know that everyone who is listening to the meeting is ready to pivot from thinking about clinical trial considerations to hearing your overview and grounding us more in what we

should be thinking about as it relates to primary mitochondrial diseases generally. So Dr. Hirano, I'm turning the stage over to you.

[00:25:30] We are not yet hearing you.

Dr. Michio Hirano: There we go. Sorry.

Susan Winckler: There you go.

Dr. Michio Hirano: Apologies for the technical issues and I had difficulty logging in, but I'm glad I'm here. Thank you, Dr. Winckler and the Reagan-Udall Foundation for putting this program together. We're going to talk about mitochondrial diseases and as Dr. Winckler mentioned, I'm going to give a background about the diversity of mitochondrial diseases. [00:26:00] And I believe I can control the slides, although it's not working. My clicker is not working.

Susan Winckler: If you click the slide once and then you should be able to go in the corner and now you should have control.

Dr. Michio Hirano: Okay. Oops. Not...

Susan Winckler: All right, we'll advance it for [00:26:30] you.

Dr. Michio Hirano: Sorry about that.

Susan Winckler: Not a problem.

Dr. Michio Hirano: Okay, maybe a [inaudible 00:26:33] problem.

Susan Winckler: That's all right. If you just say next slide, we can handle it.

Dr. Michio Hirano: Okay, next slide please. Okay, next slide. Okay, so as we all learned in high school biology, mitochondria, the powerhouses of the cell. Next image please. There you go. This is a cartoon version of the mitochondria, which resides here in a cytoplasm of the cell. And it's [00:27:00] unique in having its own DNA, the mitochondrial DNA, which is very important for these diseases. Okay, next slide. Okay, this is a cartoon version reminding us all that the mitochondria are important in generating ATP energy primarily from the carbohydrates that are metabolized pyruvate and the cytoplasm and fatty acids. And they go through the Krebs cycle. And then this final pathway, the oxidative phosphorylation pathway. And you'll notice here that there are [00:27:30] two different kinds of symbols representing the subunits of these proteins. Some are these pinkish rectangles. These are encoded by the mitochondrial DNA and the blue ovals are subunits encoded by the nuclear DNA.

So the mitochondria are the products of two genomes, and it's the interaction of these two genomes that allows the mitochondria to make ATP energy. But

when there are mutations in either genome, they can [00:28:00] often cause mitochondrial diseases. Next slide, please. Okay. This is to show you they do more than just make ATP energy. They are very dynamic organelles. They move on these microtubular rails, and you'll see here briefly around here that they come together, fuse and divide. So they are very active organelles within our cells. Next slide please. Okay. So these are some [00:28:30] of the features of mitochondrial diseases, and therefore because of the complexity of the mitochondria, the diseases that arise from this function of mitochondria are very complicated. They're complicated because, well, the mitochondria are present in virtually all cells in the body. These diseases are often clinically complex affecting multiple systems.

They're often multi-systemic disorders. The mitochondria, as I mentioned, perform multiple functions [00:29:00] beyond just making ATP energy and the mitochondria of the product of the two genomes, as I mentioned. And there are numerous, actually hundreds of individually genetically distinct mitochondrial diseases. Slide please. So this is just to emphasize the point that these mitochondrial diseases can affect virtually every organ in the body. The only cells that lack mitochondria are the red blood cells, but their precursor cells, the stem [00:29:30] cells in the bone marrow, can be affected and cause anemia. So virtually any cell in the body can be affected by mitochondrial dysfunction. These disorders frequently affect the nervous system and muscles, so we often describe them as encephalomyopathies, but visceral organs are often frequently affected as well.

We'll go to the next slide. So this is just to highlight one publication by Dr. Zolkipli-Cunningham, and our colleagues from CHOP [00:30:00] which highlights the next... Click next please. Highlights the point that these mitochondrial disease patients report an average of 16 distinct symptoms. So they truly are multi-systemic and very complicated clinical entities. Next slide. And this is just to highlight some classic mitochondrial disorders to show you the [00:30:30] diversity and complexity of these diseases. And I'll start with the disorders of the mitochondrial DNA. This is a disease called Kearns-Sayre syndrome. This is a woman highlighted here who has a progressive external ophthalmoplegia. You can see she has ptosis and droopy eyelids and she has pigmentary retinopathy, heart block, and myopathy affecting not only her extraocular muscles but her limb muscles as well.

She's sitting next to her mother and son. This is a typical disease, [00:31:00] is typically sporadic. Next slide, in contrast to these other disorders, which I'm going to highlight, which are due to mitochondrial DNA mutations. And these are maternally inherited because mitochondrial DNAs passed exclusively from mothers to their children. This is a disease called MERRF, myoclonus epilepsy. Ragged-red fibers. We use the acronym MERRF. And it's a disease clinically defined by the myoclonic epilepsy and ataxia [00:31:30] and the ragged-red fibers, which is a hallmark of many mitochondrial disorders. This is a muscle biopsy showing the normal bluish green color of muscle fibers with this modified Gomori Trichrome Stain.

But with this patient's muscle, we see there's abnormal red staining, which is a proliferation of mitochondria here along the borders of the muscle fibers. And it's a hallmark that we used very early to identify patients with mitochondrial disease. Next slide. [00:32:00] Another maternally inherited mitochondrial disorder that we frequently encounter is MELAS, mitochondrial cephalomyopathy, lactic acidosis and stroke-like episodes. And this is clinically characterized by these atypical strokes. They're metabolic strokes that affect the cortex of the brain typically, and they don't conform to the territories of large vessels. So they're not due to large occlusions of large vessels in the brain, but rather metabolic [00:32:30] strokes affecting the surface of the brain. Nevertheless, they present like strokes with sudden onset of focal neurological deficits typically in young people.

Next slide. Next slide is an image of a patient who has Leigh Syndrome. This is a syndrome. There are over 80 genetically distinct causes of Leigh Syndrome. Typically begins in infancy or early childhood, and it presents with psychomotor regression or retardation. [00:33:00] And the diagnosis is typically made on MRIs which show the lesions in the basal ganglia or brain stem or both. And it is quite a devastating disorder, progresses rapidly and unfortunately can lead to early death in many patients. Okay, next slide. Sometimes in families we see maternally inherited Leigh syndrome. The proband has severe form of this [00:33:30] devastating encephalopathy. And others will be less affected with the neuropathy, taxidermitis pigmentosa phenotype, which is also, although milder, still very severe. And you can see here progression of the pigmentary deposits here over a course of a dozen years impairing this patient's peripheral vision. So these patients share the common genotype, and I'll explain why this occurs in a moment.

And [00:34:00] yet, they have very different phenotypes. Next slide please. And then this disease, I emphasize the point that many of these mitochondrial diseases are multisystemic, but this is an exception. In this disease, Leber hereditary optic neuropathy, this is an eye disease. It causes loss of vision usually affecting men in their late teens or early adulthood. And it's usually loss of vision in one [00:34:30] eye, followed by loss of vision in the second eye caused by optic neuropathy. And this is an example of a mitochondrial disease which affects one organ or one system, the optic nerve ocular system. Next slide please. So there are these clinically very distinct disorders that can be defined by clinical criteria, Kearns Sayre with a PEO, retinopathy, and heart block, MERRF with the myoclonic epilepsy and [00:35:00] ataxia and MELAS with the strokes. NARP/MILS with Leigh syndrome or neuropathy, ataxia, retinitis, pigmentosa, and Leber's being a pure optic neuropathy. Next slide. We've learned over the last few decades that these diseases are associated typically with characteristic mitochondrial DNA mutations, single deletions causing Kearns-Sayre syndrome, a point mutation in a tRNA lysine that causes MERRF, [00:35:30] a tRNA leucine gene mutation causing 80% of patients with MELAS and NARP being generally due to one of these two mutations in the subunit six of complex V, and Leber's being due to one of these three mutations in the

mitochondrial DNA and complex I coding subunits. So there are rough clinical phenotype, genotype correlations, but [00:36:00] we'll go to the next slide.

There are now more than 270, probably more than 300 now distinct point mutations of mitochondrial DNA as well as multiple deletions, hundreds of distinct deletions of mitochondrial DNA. So there's a great genetic diversity even just looking at the mitochondrial DNA mutations. Next slide. Just acute principles of mitochondrial DNA [00:36:30] disease and mitochondrial DNA inheritance. As I mentioned, mitochondrial DNA is exclusively maternally inherited. There are hundreds or thousands of copies of mitochondrial DNA in each cell, and as a consequence, you can have variable levels of mutation. So you can have a 100% wild type in some cells, 100% mutant, but in most cases it's heteroplasmic, meaning a mixture of mutant and normal mitochondrial DNA.

And in general, [00:37:00] one needs to have a high level of mutation in a cell, more than 70% to cause dysfunction of the cell. And it's not surprising, therefore that if you have a family with variable levels of heteroplasmy, you can have different levels of severity depending on heteroplasmy of that mutation. And also depending on the mitotic segregation, the distribution of that mutation in various tissues of the cell, of the body, sorry. And the [00:37:30] threshold effect, meaning that you have to have high levels of mutation, generally over 70% to cause mitochondrial dysfunction. Next slide.

So just to give you an example of where you can have one genotype, single deletion that causes Kearns-Sayre syndrome and have different phenotypes. And in this case, as I showed you earlier, this woman has the Kearns-Sayre syndrome and it's generally sporadic, but about 4% of women with this single deletion mutation may transmit [00:38:00] it to their child and their child may have the mutation enriched in the bone marrow tissues and cause a sideroblastic anemia. As I mentioned, red blood cells don't have mitochondria, but they're precursor cells do and that leads to the severe sideroblastic anemia because of the tissue distribution of this deletion mutation. Okay. And if that child survives with transfusions [00:38:30] anemia, they may grow up to develop Kearns-Sayre syndrome or just progressive external ophthalmoplegia. Next slide.

And then this highlights the role of heteroplasmy. The patients in the family with NARP will have 70 to 90% mutation, whereas those with Leigh syndrome will generally have more than 90% mutation. Next slide please. And now just quickly highlighting that there are now hundreds of nuclear genes that have been associated with a variety [00:39:00] of mitochondrial diseases, and that number still continues to expand. Next slide, please. Okay. And this is just to highlight that again, that beyond the ATP production, there are other functions that are required by the mitochondria to maintain their integrity and to maintain their other functions. Let's go to the next slide.

And this has led to genetic classification based [00:39:30] on the particular functions and the genes associated with those functions of the mitochondria. So you can see here the large number of genes and the numerous functions that are affected by these genetic mutations. Next slide. Just to quickly highlight that a group of these disorders are due to primary nuclear DNA mutations that lead to impairment of maintenance of mitochondrial DNA, and some of these affect the mitochondrial DNA replication machinery, and some of [00:40:00] them affect the synthesis of the building blocks for mitochondrial DNA synthesis. Next slide. Next slide, please. Okay. And the most frequent mitochondrial nuclear gene disorder is a group of disorders that are due to mutations in polymerase gamma, the mitochondrial DNA polymerase, and they can cause a variety of autosomal dominant or recessive diseases ranging from [00:40:30] very early onset hepatocerebral disease, Alpers-Huttenlocher syndrome to adult ataxia syndromes.

Next slide. So this is a heterogeneity not only due to mitochondrial DNA mutations, but also nuclear gene defects can be quite heterogeneous as well. Next slide. And this is just to highlight that these mutations and diseases are rare, but if you add them up, they're not, they reach [00:41:00] a significant number. Perhaps 70,000 patients in the United States are affected by mitochondrial disease and even higher number, one in 200 carries a mitochondrial DNA pathogenic variant. So these are not insignificant numbers. Next slide, please. And I just want to highlight that there's a huge unmet need for therapies for mitochondrial diseases. As I highlighted, these are often devastating disorders and [00:41:30] often fatal, and yet the number of FDA-approved drugs for primary mitochondrial disease, next. Is zero. So this is a huge, we need to work on this. So I hope to hear and learn from the other speakers later today. Thank you.

Primary Mitochondrial Disease Drug Development

Reenie McCarthy, JD, Stealth BioTherapeutics

Susan Winckler: Excellent, Dr. Hirano. Thank you for that overview of primary mitochondrial diseases. We have all learned a significant amount and are now thinking how [00:42:00] that applies to what we learned about clinical trial considerations. And now we want to round out with one more presentation this morning for this component and to think about drug development, therapeutic development in this space. So we are going to turn to Reenie McCarthy, who is the CEO of Stealth BioTherapeutics. And Reenie, we are, I think we might all have a sense of the challenge that you face based on our first two presentations, but [00:42:30] we'd love to hear directly from you. I'm going to step away so that we can see and hear what you have to present to us this morning.

Reenie McCarthy: Thank you so much, Dr. Winckler, thank you to Reagan Udall for inviting me today. And you're absolutely going to be hearing an echo chamber. I think this is a situation where you have the FDA, KOLs like Dr. Hirano and industry definitely singing from the same songbook in terms of some of the challenges we face in

developing [00:43:00] therapies for primary mitochondrial diseases, as well as just the severe unmet need that persists in this patient population. So for my key talking points today, I'm going to focus a little bit on aspects of trial design. This is what Dr. Choe's slides referred to as lumping or splitting. We think of this as a basket trial or essentially a non-basket trial. And I would say clinical disease presentation may influence the decision of whether to combine genotypes or divide [00:43:30] them. The mechanism of action and ADME properties of the drug in development are also relevant. In either case, frankly, whether you're doing a basket trial design or as Dr. Hirano pointed out, whether you're focusing on a single genotype, these diseases are highly heterogeneous.

So trial enrichment strategies to reduce heterogeneity are important even more so when you're looking at a basket of different mutations. Targeting common phenotypes [00:44:00] across diverse genotypes is highly dependent on investigator rigor in trial recruitment, expert consensus on classifications, and natural history characterizations. And I think these are all areas where we keep trying to work harder, but challenges have persisted. Multiple approaches, including prescriptive parameters to normalize the degree of impairment such as min-max on endpoints, reduced variability of the genotypic spectrum and optimize identification of [00:44:30] target organ systems have all had limited practical success historically. One key takeaway for us in our efforts was FDA advice to balance basket trial populations by genotype was absolutely critical in allowing detection of signals in subgroups. And from a conclusion and future considerations perspective, I'll talk about our perspective on using PROs which may more holistically measure patient benefit.

So with that, [00:45:00] at Stealth BioTherapeutics, we have been working on developing therapies for primary mitochondrial diseases for over a dozen years. We have over a dozen peer-reviewed publications reporting on results of trials we've conducted as well as case study examples for patients on drug. In the setting of primary mitochondrial diseases, Elamipretide has been assessed in the setting of Barth syndrome with an NDA submitted. Leber's hereditary optic neuropathy, where we aligned with FDA on a phase three [00:45:30] protocol but did not proceed due to trial design considerations and primary mitochondrial myopathy where we've conducted a number of trials. We're also looking at Elamipretide in secondary mitochondrial diseases. We're in phase three in dry age related macular degeneration. And our next generation compound Bevemipretide, which has actually improved transport to the brain, is entering clinical trials this year.

So with that, the question of basketting [00:46:00] or not basketting, what is the influence of a clinical disease presentation on trial design? So in the setting of Barth syndrome, a couple things. It's an ultra-rare lethal pediatric cardiovascular disease. It's caused by mutations in the TEFAZZIN gene, which lead to really severe levels deficit in cardiolipin. That is the target of our drug. So right away, you're kind of setting this apart from other mitochondrial diseases based a bit on mechanism of action considerations. It also has a very consistent [00:46:30] disease presentation. It does undulate by age, so the babies tend to have acute

cardiac distress. 50% of deaths by age one. Middle childhood is called a honeymoon period from a cardiac perspective. And then you see re-emerging cardiomyopathy and progressive myopathy starting in the early adolescent years. And so for us, the decision was made not to basket based on both that mechanistic rationale and a pretty homogenous clinical disease presentation where maybe you focus [00:47:00] on age groups but you don't try to involve other patient populations.

Lerber's also, we made a decision not to combine this, and there are several mutations, as Dr. Hirano said, that lead to Lerber's hereditary optic neuropathy. The challenge that we had in considering a basket design here, which FDA did encourage, was a basket design is that we know there's a higher incidence of spontaneous recovery with certain mutations, and we thought that would introduce too much variability and really [00:47:30] be hard to stratify since some of these are so rare in this situation. We didn't move forward into phase three due to another concern we had, which was the variable use of IDEBENONE, which is a supplement in the US, but an approved agent in Europe, and we didn't think we would be able to control for that.

Mechanism of action also has other considerations for trial design. I'm going to give you the example of PolG, which is the most common nuclear DNA mutation leading [00:48:00] to primary mitochondrial disease, as you've heard from Dr. Hirano. With Elamipretide, we see therapeutic concentrations in peripheral tissues including in the retina, but we don't see therapeutic concentrations in the brain. And when you think about a disease like PolG, the pediatric presentation tends to be more neurological in nature, and although neurological impact can lead to myopathy, if the driving factor in that is the brain, we don't think the drug [00:48:30] would get to the brain, not the appropriate patient population to study. Adolescent and early onset PolG, early adult onset PolG disorders are often driven by peripheral neuropathy or, and they can lead to ataxia. There are myopathic components to these. Elamipretide can reach peripheral tissues. It may have an effect on peripheral neuropathy. Unclear that six minute walk test would be the best endpoint for that.

So that leads to endpoint considerations. [00:49:00] Late onset disease typically is mitochondrial myopathy due to skeletal muscle dysfunction, but also can include peripheral neuropathy. So again, some of this leads to end point considerations for us. Primary mitochondrial myopathy, the guidance for this first came out in 2016. It was a new clinical consensus. Dr. Hirano was a big part of this as were many of the other experts on the phone. It was really defining primary mitochondrial [00:49:30] myopathy based on a clinical recommendation to diagnose diseases based on clinical presentation. And the key feature of primary mitochondrial myopathy is that it's going to affect predominantly, although not exclusively the skeletal muscle system. If the brain is involved, it should not be considered PMM per the definition of NORD. This opened the door to basket trial designs, but I will note that this approach has not been accepted based on our latest interactions with the European [00:50:00] regulatory agency EMA.

So this new consensus solved some development challenges and it did introduce others. The basket approach allows you to recruit trials more rapidly, right? There's more patients available. Some of these specific genotypes and sub disorders or ultra-rare diseases where recruitment is really challenging. It informed a functional end point selection. If the disease is primarily [00:50:30] affecting the skeletal muscle, end points like six minute walk, five times, sit to stand, triple timed up and go are all relevant functionally, but it is highly reliant on clinician identification of the predominance of the skeletal muscle myopathy. It does require enrichment via inclusion, exclusion criteria to make sure the degree of impairment is well balanced. Notably, it allows detection of specific responsive genotypes, which was huge for us in the setting of primary mitochondrial myopathy in [00:51:00] identifying PolG as a potential target patient population.

So we've been at this for a dozen years. We started in 2014, really just in the setting of mitochondrial myopathy running trials. Our Leber's and our Barth program started around the same time. There have been others who have come and go in that timeframe, and I'm thrilled to see other companies joining today who are still in the fight alongside us. So you see Reata's MOTOR trial enrolling both mitochondrial and nuclear [00:51:30] DNA mutations. Reneo focused on mitochondrial DNA. Astellas was focusing on both mitochondrial and nuclear DNA.

So one last hammering home of some of the differences, looking at PolG, which is the largest group of nuclear DNA mutations contributing to PMM. Just looking at these different syndromes, what you're going to see is brain, muscle, peripheral nervous system, three separate organ systems contributing to disease presentation. [00:52:00] You're going to see the same thing in the setting of the 3243 mitochondrial DNA mutations or MT-TL1. That was the largest group of mitochondrial DNA mutations we've seen in our PMM trials. And again, high prevalence of neurological sequelae, psychological in some cases as well as myopathy.

So targeting skeletal muscle myopathy across diverse genotypes, again becomes highly dependent on picking the right patients, [00:52:30] identifying that it's the skeletal muscle system that's contributing to the myopathy. Our first trial was MMPOWER. That was a dose ranging trial. We enrolled 36 patients across three doses, and placebo. Saw a significant improvement in the six walk test in the highest dose cohort. Across our trials mild to moderate injection site reactions are the most common adverse finding. So this brought us into phase two. We enrolled [00:53:00] the same patients from the phase one into phase two. Before we started phase two, we developed a patient-focused, patient-reported outcome assessment, really following FDA guidance to interview patients and identify the most problematic symptoms of the disease. We identified 10 symptoms that patients identified as contributing to their disease, and we also looked for patients to identify their most bothersome symptom [00:53:30] at baseline so that we could track progress on that.

Again, the phase two clinical trial enrolled the same patients. It was a double-blind placebo-controlled crossover trial. We saw a trend that was not significant on the six-minute walk test. We saw signals of a carryover effect, which is a consideration in crossover trials, and we saw a highly significant reduction in fatigue as well as an improvement in the novel most bothersome symptom measurement. From there, [00:54:00] we went into phase three and this was a basket trial design. We enrolled 218 subjects with both nuclear and mitochondrial DNA mutations. FDA strongly encouraged us to stratify this trial so that we could detect signals in subgroups, and we did indeed detect a signal in a subgroup. So about 73% of the trial was mitochondrial DNA mutations, most commonly MT-TL mutations and about in [00:54:30] the nuclear DNA mutations, the most common genotype was PolG. We used an adjudication committee to confirm that these were pathogenic variants associated with myopathy.

The challenges we found with this basket approach, the trial was not successful, included that in the mitochondrial DNA mutation group, we saw highly variable heteroplasmy. There was an imbalance in heteroplasmy, and it contributed to a large placebo response [00:55:00] that we observed in patients, particularly with MT-TL1 mutations. We've been told that placebo response, so we observed heteroplasmy as a contributor. KOLs have also told us that in MT-TL1, as well as in Twinkle a larger than expected placebo response can also be potentially risky because of the neuropsych sequelae of these diseases. Patients with nuclear DNA mutations randomized to Elamipretide demonstrated nominally [00:55:30] significant improvement on the six-minute walk test relative to placebo with a significant exposure response relationship. So we brought these learnings forward into a phase three clinical trial, really focused on patients with nuclear DNA mutations, and again, among that nuclear subgroup from our MMPOWER-3 trial, PolG was the best-represented mutation and significant improvements on six-minute walk.

Our NuPOWER trial, again [00:56:00] limited to nuclear DNA mutations. Primary endpoint still six-minute walk because that's how we could power it based on prior learnings. But we also looked at patient global impression and the most bothersome symptom. So tried to really enrich this for the target patients focusing on nuclear DNA mutations, numerous enrichments to ensure myopathic phenotype, including known comorbidities, and really instructing our investigators to rule out [00:56:30] non-skeletal muscle contribution to myopathy, increase the dose due to the exposure-response relationship, increase the treatment period due to learnings from Barth syndrome that nine months or more is best. We have not read out on the data from the NuPOWER trial. We are expecting full data analysis by year end. We did enroll 102 subjects. Most had replisome-related mutations. Most of these, once again, a majority were mutations in PolG-1.

[00:57:00] The question we have based on baseline inclusion criteria and the baseline demographics that we've observed are whether we were successful in enriching for the myopathic phenotype. The most bothersome symptom was

variable, about 20% identifying non-myopathic symptoms as most bothersome. Medical history did confirm that patients with ataxic and neuropathic presentation were included. And we saw variable age of onset [00:57:30] in the medical history, which we know corresponds to potentially a more neurological or peripheral neuropathic phenotype.

We did some other work to characterize what's most important to patients. This was in the setting of Barth syndrome where we were able to do semi-structured qualitative interviews of patients after they finished the trial, but while they were not aware of how they were randomized during the trial. FDA actually came out with guidance a couple of years [00:58:00] after we did this. Super informative in terms of symptoms that weren't on the radar screen of either US or the patient community. Several patients reporting cessation of nighttime bedwetting, improvement in appetite, the concept of recovery from activities. So again, this was really informative and kind of goes into some of our thinking around the utility of the most bothersome symptom construct.

In terms of our thoughts on ramifications for outcome assessments, [00:58:30] one of the things, if we are not successful really balancing the heterogeneity or screening out for it through inclusion criteria or identification of common phenotypes, should we look at more global impressions of disease severity so that patients can be making a holistic assessment of whether they've improved during the trial? So the patient global impression of disease severity is one that FDA is increasingly recommending in clinical trials. Again, our most bothersome symptom construct [00:59:00] was intended to also give a very individualized assessment of how patients are doing. Intriguing for us is the idea of coupling that with a functional endpoint that would be most predictive based on that most bothersome symptom. So if that's muscle weakness, for example, perhaps six-minute walk is best. If it's balance that's most problematic, maybe you should be looking at some sort of a balance scale.

So again, takeaways. [00:59:30] Trial design should be informed by the agent in development as well as the clinical disease presentation. I would say that reducing heterogeneity has been challenging from a historical perspective. Is important to try to stratify or balance if you can, to detect signals or potentially consider endpoints that are going to allow more individualized assessment of benefit from patients irrespective of what's driving that benefit. I think one thing we can all agree on as [01:00:00] though the challenges are great, the unmet need here is tremendous and patients are waiting. So I applaud the FDA again, as well as all of the KOLs and other industry partners for continuing to work on this. It's super important. Thanks for your time.

Reactor Panel

Anna Choe, MD, MPH, Center for Drug Evaluation and Research, FDA
Marni Falk, MD, Children's Hospital of Philadelphia & University of Pennsylvania Perelman School of Medicine
Michio Hirano, MD, Columbia University Medical Center

Reenie McCarthy, JD, Stealth BioTherapeutics
Brian Tseng, MD, PhD, The POLG Foundation
Philip Yeske, PhD, United Mitochondrial Disease Foundation

Susan Winckler: I am thinking, so I'll note here, we're going to go to our reactive panelists, but Dr. McCarthy, that was fabulous and helping us round out and think through [01:00:30] as we had thought about the clinical presentation. And then we go to the how it is, what the regulatory expectation is, what does good look like, and then you helped us navigate and think through okay with that, what does good look like, how very challenging it might be to get to that good and continuing to do that. So really appreciate it.

We're bringing back up on screen here our other two speakers [01:01:00] from this session, so Dr. Choe and Dr. Hirano, and we are welcoming to our virtual stage three reactor panelists. They are going to help us reflect on and explore the presentations that we just heard. So I welcome Dr. Marni Falk from the Children's Hospital of Philadelphia and University of Pennsylvania, Perelman School of Medicine. I welcome also Dr. Brian Tseng from The POLG Foundation and Dr. Philip Yeske from the [01:01:30] United Mitochondrial Disease Foundation. So I'll note to our reactor panelists, you get to go first. Do you have any questions for our presenters or are you thinking and are struck by anything? I was going to say, Brian, that you unmuted first, but I actually can't tell. Dr. Yeske might have, but I already called on you Dr. Tseng, so you go first. Then I'm going to [01:02:00] turn to Dr. Yeske, and then I'll pick up Dr. Falk. But let's start that discussion.

Dr. Brian Tseng: Thank you Dr. Winckler. Hope everyone can hear me okay.

Susan Winckler: We can.

Dr. Brian Tseng: What an amazing morning. Three, I don't know, a triple crown of Dr. Choe, Dr. Hirano, and Reenie. It was fabulous. I picked up on a number of things that I wish I could dive in deep on, but I just want to come back to the idea that this is all about for patients. And the question that I really struggle with sometimes when I've had interactions [01:02:30] with the FDA and other agencies is on endpoint selection when we say the effect must be clinically meaningful, and we have heard before Dr. Woodcock say how a patient feels, functions, or survives, but in ultra-rare, rare space, the example Dr. Choe, you gave of one of the recent approvals, I was struck that liver spleen size was part of the approval. And [01:03:00] when I think about does a patient feel that, is that a function, it clearly predicts health and the regression of a size.

But I'd like to hear each of you three speakers take on clinically meaningful because it seems like we all seem to know what it is. Is that because the doctor says so? Is it because the patient says so? And to Reenie your point of importance to PROs. But I would love to hear people really try to help me disentangle clinically meaningful.

Susan Winckler: That's quite [01:03:30] the opener. Who of our speakers wants to jump in? Well, Reenie said first, so-

Reenie McCarthy: I'll jump in just really where that was super helpful for us. Again, I'll go to Barth syndrome because we were able to do semi-structured qualitative interviews. And so patients contextualizing, like we saw improvements in muscle strength. They were durable, they were consistent. And patients saying, "Well, I used to have to use two hands to pick up a half gallon of milk, and now I can carry one in both hands, or I can carry two bags of potatoes. I [01:04:00] can help my mom unload the groceries now." Recovery from activities, kids saying, "Well, I could always walk to the stop sign and back, but I had to rest longer in between doing it and then I would have to sleep for two hours." So if that's only 10 minutes, that's a big difference. It does help contextualize clinical meaningfulness. I'm not sure you get it from the endpoints themselves.

Susan Winckler: Dr. Hirano, please.

Dr. Michio Hirano: Yes. I just want to make two points about what's clinically meaningful. Of course, [01:04:30] there are some symptoms that the patients report, like headaches or fatigue that only the patient can describe. So you have to depend on patient-reported outcomes there, whereas, but I think a general issue that I bring up over and over with my colleagues, is that I think we really need therapies that have big effects for these rare diseases because they're so devastating so that it's obvious that it's having an effect. It's nice to have a minor change in [01:05:00] let's say a liver size, as you said, but if you could change a person from being on a ventilator to coming off the ventilator or wheelchair-bound to becoming able to walk, those are huge effects. And I don't think anyone will dispute that those are clinically meaningful.

Susan Winckler: Yeah. Thanks Dr. Hirano. Dr. Choe, do you want to add anything to that?

Dr. Anna Choe: Sure. I don't think there's anything new necessarily, but I do think it's a really good question. And that's something that I struggled with as a newer reviewer when I first joined the [01:05:30] FDA, trying to understand what does this really mean? And I think Dr. Hirano really put us on a point that it depends on multiple factors. It's not just on the endpoint itself. It also depends on the treatment effect size. That example that you refer to, I think there's some precedent that informs that. I think spleen site liver size typically would not be considered clinically meaningful endpoints, but I think that it really depends on indication of the precedent as well.

And to Reenie McCarthy's point, [01:06:00] we are really encouraging that sponsors conduct qualitative studies and incorporate exit interviews so that we are able to create that evidence to support the meaningfulness interpretation. I think especially for COA endpoints that tend to have more smaller treatment effect that we see towards the end of the trial, I think that can be really useful. The functional endpoints a lot of times, I think can be sort of more self-evident in terms of the clinical meaningfulness because it relates well to how it affects

[01:06:30] how patient function. So it depends on endpoints, depends on the treatment effect, multiple factors that go into consideration.

Susan Winckler: Yeah. Great. Thank you all. Dr. Yeske, I'll turn to you.

Dr. Philip Yeske: Thanks so much Dr. Winckler, and again, a big thank you to the speakers this morning. It was a great start to this workshop. Much like Dr. Tseng, I have about three or four pages of notes already in a lot of different directions that we could go with it. But [01:07:00] I really appreciated hearing through all of the speakers, that all stakeholders need to be involved in singing from the same songbook, I think Reenie you used this phrase. So critical, right? That the FDA, clinicians, industry and patients and patient advocacy groups, I think really important. So standardization is sort of one topic perhaps for the [01:07:30] speakers to speak to.

But in thinking about how all of these stakeholders involved, one aspect that kind of came up over and over is natural history study, right? Natural history study data and how important it is. So I know we have limited time today, and this is one we could speak for hours on, but perhaps each of the speakers could comment on what do we have, how can we collect it better? [01:08:00] And importantly, for industry and Reenie for you particularly, the absence of natural history data, how that impacted your clinical development plan. What can we learn from that? So maybe Dr. Choe, or just kind of speaking of where we are with natural history data.

Dr. Anna Choe: Sure. I'm probably not the best person to comment on what we know about the natural history data. I'm actually looking forward to learning more about it today as well. [01:08:30] But in general, natural history data does help us define the population, but also the study design. I think identifying prognostic variables that can inform the population selection has been really useful for us. It can also help us inform the endpoint selection. We just haven't seen a lot of natural histories that we have reviewed that clearly defines even six-minute walk tests and PMM. Some of the studies I've looked at, it really varied. There wasn't [01:09:00] a clear course that was documented that made it really challenging. So I don't know what the answer is. I think it depends on the disease and what we understand. I would say more data that we have to inform the trial, the better off we will be.

Dr. Philip Yeske: Yeah, I think one of the takeaways, right, you elucidated trial design, population selection, and endpoint selection. Of course that matters in interventional trials, but it's also very important, right, for natural history [01:09:30] studies, and that's where we can refine those things to make sure the transition then to interventional arms goes even more smoothly. But Dr. Hirano I know you're in a position to talk about where we're at with natural history data and how do we do it better going forward?

Dr. Michio Hirano: Well, I think we really need to begin with registries to enroll patients into mitoSHARE and the NAMDC registries because we need to start with that and

then we engage the patients. And then over time, serially we'll be able to track patients. [01:10:00] So of course it sounds easy, but it's not because we don't know exactly which outcome measures. And I think that'll be discussed later, which outcome measures are best for which patients and it takes money. And frankly, we don't have a lot of money for natural history studies. An alternative that we initially tried with another mitochondrial disease TK-II deficiency, was to do it retroactively, retrospectively where we, because [01:10:30] it's a myopathy, there were clear milestones, motor milestones that we could ask patients when do they lose the ability to walk or sit or breathe independently. So that was very helpful in setting a rough guideline of what the natural history might be and even survival, of course. So I think that we can try to collect some information retrospectively, but to do it right prospectively we'll need more resources and greater efforts and hopefully some guidance later today [01:11:00] about outcome measures.

Dr. Philip Yeske: And that really is about maximizing the value of every data point that's ever been collected retrospectively, but also then going forward prospectively. I would also underscore the importance of certain key infrastructure that's necessary to be successful with this. You mentioned patient registries, but we know biorepositories, genomic repositories, all of these clinical research networks are really critical, and [01:11:30] UMDF and many of its partners and other PACs have spent the past decades building up this infrastructure. I, for sure, would want to leave with the recommendation that all stakeholders leverage what we've built, what we have to take advantage of this infrastructure in a standardized way to try to advance these natural history studies. Reenie, I guess that leaves you to talk about the context of a clinical development [01:12:00] plan for industry sponsor.

Reenie McCarthy: Thanks, Phil. It's a great question. I think in the setting of Barth Syndrome, which is a little bit more discreet, and so we did a retrospective natural history control trial. I'm going to use that example. Barth Syndrome Foundation partnered with Kennedy Krieger Institute who actually came to some of the biannual patient conferences and conducted clinical assessments. And with that, they were able to identify in a longitudinal fashion what endpoints change over time, so [01:12:30] what endpoints progress and what endpoints don't. So that was actually super helpful. They looked at this relative to values for normative age match controls so they could tell where Barth Syndrome patients were impaired on six-minute walk, on muscle strength, on balance, whether the assessments were sensitive enough relative to those normative match controls and importantly progression over time. So it's a small disease where they were able to do that, but it gave you longitudinal data, it gave you longitudinal data [01:13:00] on clinical trial endpoints so it was really informative for drug development.

The challenge, I think, comes in using that data. Again, Barth is ultra rare, so as a control to try to use natural history data as a control, if you listen to our adcom, lots of questions on how you prognostically match or extrapolate data from natural history data sets. And so I think there continues to be some question

about the utility of natural history data for effort- [01:13:30] dependent endpoints as a control. Really, really informative for trial design, I think we need to work towards a better understanding of how we can use it to solve for some of the recruitment and again, control challenges we have with ultra rare.

Susan Winckler: Yeah, I am going to turn to Dr. Falk and then I know I can sense that our reactor panelists have more that they want to ask about, but we're going to go in turn at least here. Dr. Falk to you.

Dr. Marni Falk: Yeah, that was wonderful. Thank you to [01:14:00] the panelists and of course to my reactor panelist, colleagues, great discussion. We certainly heard a lot about the variability, which is always the starting point when you speak about energy diseases of the mitochondria. And we try to force, I think, a lot of differences into some commonality. And I think Reemie, you were commenting, and of course, Michio, that it's a multi-system disease and that you've tried to go after patient-important symptoms. [01:14:30] I think that's something we should really dig into a little bit more in terms of what we could be doing differently rather than always doing the same thing that doesn't work. The definition of insanity is keep doing what doesn't work. And so I guess my question is how could we do things differently from your perspective?

So for example, there's mobile outcomes that can maybe capture some of what Reemie you were subjectively describing. There's variability in terms of safety risks for some of these therapies. How is that accounting for the degree of benefit? And I think back [01:15:00] to Brian's question, who determines how much benefit there is? And we haven't really spoken about how to think about N of 1 outcomes. So patient-important symptoms and matched function and then assessing whether there really is an objective benefit and setting that up front. So I'd like to talk about some of that a little bit. And maybe also, if you could sprinkle in there, are we judging our outcomes properly? So of course there's the primary outcome, but we've learned that a lot of therapies, [01:15:30] including some that were shared with us today by Dr. Choe, maybe didn't even meet their primary outcome. So how can we better understand when there's a benefit, and I'm just going to throw out there AI, are there opportunities for different ways to really integrate the data and think about what is the relative safety and benefit of what we've learned to really inform the decision-making?

Susan Winckler: I'll just say as you get ready to jump in, Dr. Falk, it says, "If you were channeling of our attendees who had just shared [01:16:00] something similar in saying, so what's the application of AI here?" So Reenie, would you like to go first on this one?

Reenie McCarthy: Sure. So a lot of parts to that, Marni. I think that the challenge for us, if you're trying to enrich by an organ system involvement, and we've talked about this with you and other experts, it can be sometimes hard to know what organ system is contributing to the dysfunction. It's probably a combination, quite frankly. [01:16:30] And again, with a drug like ours where we think there's broad

distribution, but not to the brain, that's important in these diseases. And so I'm not sure how to do that better. I think that that's for you and other experts to actually help us understand because we understand it's hard even for investigators to make those judgment calls.

Mobile outcomes, very intriguing. As you know, we've looked at accelerometry both in Barth Syndrome and in PMM. The data remains really [01:17:00] noisy and hard to interpret, and I think from a practical perspective, and this is for Dr. Choe, how valuable is it even for FDA because it adds a lot of burden and expense to a trial and it's tough to interpret, but the promise of it is huge because then you know what happens outside the clinic. Something we found really interesting in looking at our Barth Syndrome data, we're doing some continuing to look at this in PMM, is, don't just look at fatigue overall, [01:17:30] but look at fatigue after a trial visit. Does it spike or does it go down? Because then that's maybe a more important time to capture fatigue for patients. So I think even looking at some of the data different ways becomes really important.

On the individual benefit and how that pairs to functional. For us, we're intrigued by that concept and it could be not how we designed it, but others, if patients identify their most problematic symptom at the beginning of the trial, could you then pair that with one of the functional assessments to [01:18:00] wait, the importance of it on a more individualized basis. So those are my top line thoughts.

Susan Winckler: Thank you. Dr. Hirano, do you want to go? Oh, Anna, un-muted. Dr. Choe, you can go next and then we'll turn to Dr. Hirano.

Dr. Anna Choe: Sure. Thank you. Dr. Falk, I think you included a lot of topics, so I try to jot down some notes, but I may not touch on everything. I guess some of my initial reactions when I was hearing your question was I think Dr. Knoble [01:18:30] is going to touch on this. I would love to see more diverse outcomes, not necessarily diverse concepts that we're measuring. If we don't really know what we're targeting, not just looking at myopathy, but looking at diverse outcomes. Thinking about exploratory trials, we may not see a signal necessarily in PMD for the shorter trial because we typically need longer duration. But there's some other utility. We may see some signals we may be able to improve our measurement tools. I think I'm [01:19:00] always thinking about how do we reduce variability? And to Reenie McCarthy's point about mobile outcomes, I presume you're talking about digital health measurements. That's not the okay. Yeah, and I think-

Reenie McCarthy: Accelerometry is what we looked at, yeah.

Dr. Anna Choe: I think the noise is an issue and having more data. I think when we're thinking about trial design, we think, well, having more data is great, but I think it ends up prone to more error, as well. So depending on the purpose of the trial for exploratory trial, sure, having more data, we can explore [01:19:30] more signals. But for the pivotal trial, I think ensuring data quality is something that

we need to think about balancing, and we're also thinking about digital health technologies for endpoints. What I would love to see is more data to inform the pivotal trials, whether it's exploratory trials, data from non-clinical models. I know it's challenging this space because of the disease mechanism, pathophysiology and then the drug mechanism. So it's not the same, but that is [01:20:00] on my wishlist moving forward.

Dr. Marni Falk: And what about the AI part of it? Other ways to analyze? I know you have a brilliant team of reviewers that come on and review every point. Can you articulate for us how that might change or really incorporate new algorithms for determining if there is or isn't a benefit or a harm?

Dr. Anna Choe: Yeah, I think that's an area that our agency is interested in. There's been part, there's been community of practice for artificial [01:20:30] intelligence. There's been a lot of case studies. People are piloting things within the agency, but it's not something that's widely implemented, yet. I think safety comes first. So with new technologies, we want to make sure that it's working before we implement it widely. So currently we're using what we know works, but it's not something that we're opposed to implementing so that we can increase efficiency and [01:21:00] do our job better. So yeah, I'm really interested to see how the general public develops AI and how that influences our work and how we can do our job better. But at this point, I think we have to think about the benefit risk, and currently, we want to ensure safety before trying innovative technologies.

Susan Winckler: Continuing to learn.

Dr. Marni Falk: Yeah, I definitely agree. And I think one of the things that always sticks with us is in any trial, there's certain people that said, "I did so much better," [01:21:30] the subjective. And so using AI to understand what we can't see, but where the data's showing that would be really, I think, informative. Maybe not to get approval on that trial, but to guide the design for the next or maybe approval. But again, I'm glad to hear that the agency is considering it, so thank you.

Dr. Anna Choe: Thank you.

Susan Winckler: Dr. Hirano, round us out on this one.

Dr. Michio Hirano: Okay. So yeah, I want to share a few thoughts. I absolutely agree with you, Marni. These are important topics and the mobile [01:22:00] devices are great, I think in allowing us to actually assess patients in their home environment as opposed to artificially in a hospital. So what does a six-minute walk test mean? We don't really know as much as what they do at home and we could see how active they are at home and we see changes. So there's great potential there. I think there's certainly a place for the mobile devices when we're using patient reported or caregiver reported outcomes because it's so much easier to do it there at home and that's really helpful. But I think [01:22:30] when we're looking at mobile app devices as ways to measure things like activity, I think we

still need to anchor that to some measure that's accepted by, let's say, the FDA. So we can say, yes, we see that the patient has spends 30% more time active at home with this Fitbit or whatever the device is, and then that correlates with a change in a six-minute walk test that has been deemed clinically important or meaningful.

[01:23:00] So I think there's a lot to be done there. And then certainly applying AI in that way is also very important. I have a colleague here who's using AI to, well using videotapes of patients walking 10 meters and applying AI to that. How that will work out, we're not sure, but he's exploring that possibility. And then regarding the N of 1 studies, I think it's great to, as we treat patients, I know Expanded Access goes through [01:23:30] the Reagan-Udall Foundation. So that's an opportunity where, if we have patients that are trying a therapy on Expanded Access, where we can try to gather information to use that N of 1 information to see what is the magnitude of effect we're seeing under Expanded Access, and then can that be applied in a trial? It gives us a rough ballpark.

N of 1 is not just treating, but also stopping it. If they can develop a complication or something, they come off. What happens when they're off [01:24:00] the drug and then restart it. So you really want to collect that information when you can and take advantage of it because it could really help inform trials down the road.

Susan Winckler: So I'm going to say for I'm going to have each of our reactor panelists, you get one more observation or question, but we don't have time for all three to respond. So you get to choose your preferred respondent, and if they lean away from the microphone, another one's going to jump in. Dr. Yeske, fire away, [01:24:30] and then we'll go Dr. Falk and close out with Dr. Theng. Dr. Yeske.

Dr. Philip Yeske: Choose carefully is what you're saying, right?

Susan Winckler: Indeed.

Dr. Philip Yeske: Well, I think a natural follow up to the discussion that just took place around some of the limitations on digital health wearables, et cetera, with noise, there are other approaches like digital video assessments where again, very patient-friendly, those can be collected in a home environment and then [01:25:00] can be objectively measured for change. So that's very different than a wearable that's just generating a lot of raw data that then has to be interpreted in, perhaps, not such an objective way.

So Dr. Choe, you're going to be my selection. We know in other disease communities like Duchenne Muscular Dystrophy, these digital video assessments have been advanced. I don't believe they've been used as a primary [01:25:30] endpoint at this point, but could you just speak to the agency's position on the value of these? Because foundations like UMDF, POLG Foundation, Mito Foundation in Australia, we've invested in trying to advance

digital video assessments for specific types of mitochondrial disease like POLG and primary mitochondrial myopathy in the hope that those types of endpoints can play a role in future clinical trials. Your thoughts?

Dr. Anna Choe: Thank you for your question. And I will say this is not going to be the agency's position, [01:26:00] this will be my position, but I actually think that video assessments can be really helpful because then you are able to utilize a centralized rater, and again, you're reducing concerns for variability, having multiple assessors having different position. So I think it'll depend on what we're assessing. We'll have to reach agreement on what we're actually measuring and how we're going to demonstrate clinical meaningfulness. But video assessments as a concept I think could be really [01:26:30] valuable.

Dr. Philip Yeske: So as a reviewer, you would be open to seeing more of those types of data submitted as part of packages?

Dr. Anna Choe: Yeah.

Dr. Philip Yeske: Yep. Thank you.

Susan Winckler: Great. Thanks so much, Dr. Yeske. Dr. Falk?

Dr. Marni Falk: Yeah, again, great discussion. Grateful for everybody's thoughts and ideas. One of my passions is using data-driven preclinical models to advance what we are doing in the clinical space and how do we do this most efficiently? And [01:27:00] I think Dr. Choe, you mentioned this a little bit, and I'd like you to maybe expand, and of course I realize all the panelists have quite a lot of experience in this area. Can you explain how you utilize the confirmatory evidence? Because I think it's becoming even more important in these diseases to not just assess safety, but also some level of potential benefit. And I think we've talked about this before the idea, but well, if you wanted to start in a child, how much more evidence do you even [01:27:30] need? I know we'll be talking about this later today, but can you maybe just give us some of your thoughts and of course the others if there's time, about the role of confirmatory preclinical evidence?

Dr. Anna Choe: Yeah, so I think you're referring to non-clinical data, and then confirmatory evidence is a piece of that. So in general, non-clinical evidence, I think typically we think about it informing our safety signals based on the tox data, but we also look at the efficacy data from some of the pharmacology studies to figure out are there any signals? [01:28:00] What are some things that we can look at in terms of enrolling pediatric patients, of course, we typically require higher, more data to justify the risks that we're putting patients under. And that's where it comes in, you have to think about is there proof of concepts? Does the exposure, do we think this is going to work? Are we really underdosing that we're not going to see any benefit for kids? So that's where that comes in. And that can be derived from non-clinical models, as well.

When it comes to confirmatory evidence, that [01:28:30] is when we think about the whole substantial evidence package, and when you're using a single trial because you don't have that second study replicating that study, that we rely on CE, and that can come from a variety of sources, but non-clinical model is something that you can rely on to substantiate the findings of the first trial. I would say that the bar is not low and that I encourage you to talk about your plan with the agency [01:29:00] before you move forward with that plan. There's also a guidance on CE that's available, Substantial Evidence Effectiveness, I believe, based on a single, radical, well-controlled trial that goes through all the considerations. But your CE package can be more than one thing. It doesn't need to be just non-clinical model. It can be additional data, as well. And it really depends on your single trial. If you have a really strong trial, then your CE evidence, the strength is going to vary based [01:29:30] on that. So that's also something else to think about. Does that answer your question, Dr. Falk.

Dr. Marni Falk: It was wonderful.

Susan Winckler: Right. All right, Dr. Theng. Last question for us.

Dr. Brian Tseng: I guess I'm going to hit up Reenie first, and maybe Dr. Choe will chime in afterwards. Reenie, I have to praise Stealth for really seeking early advice and doing really strategic iterative learning as your company and your programs have progressed. [01:30:00] I suppose we know a lot of companies just abandon ship, cut bait, and run. My question for you is how would the regulatory pathway look if it was more harmonized? This may be just me being too provocative, too much of a pipe dream, but if the FDA and EMA actually agreed on what a mitochondrial patient looked like and could outcome measures be agreed upon in a master protocol, could there actually be a master effort that could pull both agencies in together?

[01:30:30] And maybe Dr. Choe, you can tell us what the latest is on those monthly calls transatlantic and other countries, but let me just throw it out to the two of you and see what you come back with, please. Because rare disease is rare.

Reenie McCarthy: It was a real conundrum for us in the setting of primary mitochondrial myopathy. I don't think we had the same concerns with Barth. LHON, we had similar concerns just because of different approved medications being used. But in PMM, EMA when we went for guidance couple times, [01:31:00] really just doesn't recognize PMM as an indication, really looking more for genotypic specific buckets. And it just poses challenges for pursuing regulatory pathways in Europe. You can still do trials there and we did, but I think it does pose some regulatory hurdles that we still aren't quite sure how to navigate, quite frankly. I think because in PMM we really were seeing signals [01:31:30] in subgroups, it became less important to us over time. But that was certainly a challenge at the outset. Dr. Choe?ee

Dr. Anna Choe: I'll just add that we do have regular calls with other global agencies to harmonize when we can. I think the issue is that our regulatory frameworks are not identical. So I think that's where we can't completely harmonize as I think sponsors or academicians may wish.

Susan Winckler: Yeah, I think Dr. Choe, I [01:32:00] just underscore in having seen the components, there's different structures as well as there can, on occasion, be different risk-benefit, just calculations culturally, that become important. Dr. Hirano, we actually do have time if there's anything you want to add to this one just before we close out.

Dr. Michio Hirano: No, I think that they've said everything. Thank you.

Susan Winckler: All right, then let me say [01:32:30] it is that moment when I say I wish we had more time and yet, we have so much more material to cover that I want to thank each of our speakers for the morning sessions and helping set the stage. And then Dr. Yeske, Falk, and Theng, you've set the standard for let's ask great questions and continue the knowledge sharing in the discussion session. So with that I'm going to take us, we're take a quick 10-minute break. [01:33:00] We've learned in some of our virtual meetings that when you just march on through, we wear everyone out. So we're going to take a quick 10-minute break. We will look forward to seeing everyone back on screen at 11:45 Eastern time.

Selecting Patient-focused Outcomes and Statistical Considerations

Naomi Knoble, PhD, Center for Drug Evaluation and Research, FDA
Yan Wang, PhD, Center for Drug Evaluation and Research, FDA

Susan Winckler: All right, let's return to our meeting. Welcome back, everyone from the break. It's time now to move from our initial foundation setting discussion in that really rich engagement [01:33:30] to turn to another component where we want to focus on selecting patient focused outcomes and statistical considerations. Our next session features two speakers from the FDA. I'm going to turn first to Dr. Naomi Knoble, who is a Pediatric Neuropsychologist and Associate Director specializing in rare disease measurement in clinical trials within FDA's Center for Drug Evaluation research. And then following Dr. Knoble, we will turn to [01:34:00] Dr. Yan Wang, who is a statistical reviewer within CDER, and she's going to expound on global tests for multiple endpoints in rare disease clinical trials. Dr. Knoble, we are going to turn to you first.

Dr. Naomi Knoble: Thanks so much Susan, and thank you all so much for this opportunity here. So as Susan indicated, Dr. Wang and I will discuss selecting patient-focused outcomes and statistical considerations here in our talk. The views that will present are ours and [01:34:30] we have no conflicts of interest to disclose.

In this brief talk on selecting patient-focused measurement, my talking points for you are focused on two key questions. How do you select the key symptoms or disease impacts and outcomes measurement, and then how do you set

endpoints up for success in clinical trials? And just in full disclosure to say that there's no perfect measurement, or at least not one that I'm aware of. [01:35:00] Every measurement decision has advantages and disadvantages. In this talk, I'll highlight some and look at some solutions to working with those advantages and disadvantages and ways to mitigate risks. And in this brief presentation, too, at times I'll be speaking about COA selection, Clinical Outcome Assessment selection, which includes the Clinical Outcome Assessment itself, which is a measure that describes or reflects how patients feel functioners survive.

We'll talk about the score, which is a separate [01:35:30] piece, but related of course to the assessment itself. And then also at times, I'll talk about the end point precisely defined variable that includes, of course, the COA score, but then is also tailored to the product's mechanism of action and the study design and more. So these are three different things. They're very closely related of course, and focusing on endpoints for just a moment here. Before getting to the key themes of this talk, I'd like to highlight here that there are various ways [01:36:00] to work with endpoints. And on this slide, which I won't go into in great detail, there are different strategies to working with endpoints. I referenced here two guidances on this topic, and Dr. Wang, in her portion of the talk, will address this as well. But I mentioned endpoints here because they involve the COA end score and the end point is another critical layer of the clinical trial design and measurement specification for measuring outcomes that matter.

I'd like to mention here, too, that assessments that are [01:36:30] used by providers in clinic aren't always suited to be the best measurement approach used in an end point that's best in a trial. And to state that another way, maybe what we do in clinic to diagnose and monitor patients doesn't always work in clinical trials. There are a lot of reasons for that. But I think the objective of a clinical trial is to design a study to detect a treatment effective, one exists, for a new treatment, and then to show that key disease features have improved, [01:37:00] although maybe not all of those disease features. So it may not always be reasonable for one single endpoint to include all comprehensive symptoms and especially in a heterogeneous disease like PMD. And so the entirety of all of the endpoints working together can measure what's necessary to convey clinical benefit of a new treatment.

And so to address the two key questions in this talk, the first is how to select the best [01:37:30] symptom and outcome measures. The second is how to set endpoints up for success in clinical trials. I have four recommendations here. One is to engage patients, caregivers, clinical experts and others, and use their feedback and measurement approach. The second is to get specific and identify specific aspects of targeted clinical outcomes. The third is what I call measurement diversity. And that's selecting lots of types of COAs and other ways of measurement to offer [01:38:00] distinct sources of evidence within the overall endpoint strategy. And then last is what I call test-driving. You could also call this pilot testing or conducting an observational study. And this is a way to

learn about measurements, learn about endpoints before using them in pivotal trials. And these four points are largely derived from our patient-focused drug development guidance series, which many of you may already be familiar with.

And in fact, as I was preparing [01:38:30] for this talk, PMD clinical experts said, "You know, Naomi, our community is already pretty fluent with what is a COA and what is fit for purpose." For those of you watching today, if you are not familiar with these terms, I refer you to the guidances that are listed on this slide here. And just very briefly, when we use the term fit for purpose, it's a regulatory determination that a clinical outcome assessment that is used in an endpoint is suited, it's valid [01:39:00] and reliable, it's meant for the context of use of the disease, the study duration, the study design, the mechanism of action, and other factors here.

So turning to my four recommendations, there's engage, get specific, measurement diversity, and test drive. And these are my high level summary points of this COA roadmap. And so in this roadmap, it may appear to be a linear process that moves from the left side to the right side. However, in my experience, it [01:39:30] is iterative reasonable outcomes and COAs may be selected only to find that they just don't perform well. And maybe it's the medical product that's tested, maybe it's a study design, or maybe it's that the outcome wasn't specific enough. Maybe it wasn't evaluated before the trial. But when this process becomes iterative, it's not because the information gathered the first time was wrong. It's often just because there's a need for refinement, like getting more specific or trying alternative measurement [01:40:00] approaches.

And so to our first point of engagement here - oh, sorry, there we go. To our first point of engagement, the mito community has engaged and has created a very strong foundation of understanding what matters to patients. And so we take this information and then ask the question, how do we best select the symptoms to focus on? So after engagement [01:40:30] is getting specific about what we're measuring. And so in getting specific we ask what aspects or attributes of the symptom or the impact should we be focusing on here? Is it severity, frequency, duration? I've listed a few here on this slide. Duration can be invaluable and I'm just going to focus on duration and severity because of time constraints. But duration can be invaluable, especially for symptoms [01:41:00] that fluctuate. And sometimes it's measured best by a daily score. You can think of the duration of a headache and you want to prove that you've shortened the duration of a headache as your treatment outcome, for example.

Symptom severity can be measured in lots of ways. I'm partial to a daily diary approach, as well. A daily diary approach for symptom severity, you can use it across diseases and conditions to measure lots of things, of course. Reenie McCarthy actually mentioned this in the first session with an example [01:41:30] of measuring exercise and tolerance. And so with a daily diary example, you could measure the day after exertion to see if you've truncated or shortened

the recovery time. So a daily diary is a nice way to do this and look at, again, that question of severity or duration.

So in the following two slides, I have some examples of ways that symptom severity have been measured with daily diaries. And I want to ask you all to use the spirit of mind [01:42:00] of what we like to call transferable skills. This spirit, the approach to measurement could be transferred to PMD even though it may not have been used in the PMD space yet. So transferable skills, it's like when you're looking for a new job and you're like, well, how can I move these skills to this new position that I want? And that's the mindset I'm asking you to hold as I go through the next example or two.

So here, I've listed daily diary examples for clinically relevant [01:42:30] symptom severity measurements. So in the first example, we have reduction in monthly headache days or monthly migraine days. We have reduction in weekly mean daily maximal hunger score. And these examples that I've listed here, I've listed some approval examples that have used them, as well. I won't go through them point by point, but this is to illustrate that you can use daily diaries to measure these very specific aspects of selected symptoms that are important and relevant to patients. In PMD, I think there's [01:43:00] value in exploring using single items in a daily diary, maybe to evaluate a reduction in maximum symptom severity, like reduction in maximum fatigue severity in a week, for example. And the data shows that patients are really good partners in collecting evidence in this way, especially when the daily diaries are brief, clear, and the reason for them is explained to patients. There are disadvantages and they include a patient burden, especially if the diaries are long. It also includes data integrity issues, [01:43:30] but I think that risk can be mitigated with test-driving your measurement and pilot testing it before using it.

And so I have an example of this for you on the next slide. And this example on the left side is in a disease called Progressive Familial Intrahepatic Cholestasis, and this is a very complex disease. You can see at the bottom of this graph, itching was reported by patients and caregivers as one of the most interfering symptoms. And so the drug [01:44:00] developers who had already engaged with patients got really specific about measuring itching. And I have a synopsis here, although you can also find this case example is publicly available on our Accelerating Rare Disease Cures Leader 3-D page. And you see here that they selected caregiver reported daily diaries of observed scratching. Many of the children, many of the people who are in this clinical trial were children so caregivers were the best reporters here. And they looked at a reduction [01:44:30] of at least one point in the patient's worst weekly average scratching score. And so that's an example of engagement and then also getting very specific about what's being measured.

Turning next to measurement diversity. So this is addressing the question of how do we set endpoints up for success in clinical trials? And I think you've probably heard the financial advice diversify your portfolio. I think the same advice can apply to clinical trials, [01:45:00] diversify your measurement

portfolio. I think in the PMD space there's been creative, interesting scientific exploration into measurement. We heard this in the first session from Dr. Yasky and others. The point is, again, a perfect measurement may not exist and most measures have one or several limitations. So if you're just using one measurement for all of the evidence in a clinical trial, its limitations [01:45:30] might be amplified and it can undermine the evidence. And so diversifying your measurement within the endpoint hierarchy or at least endpoint strategy, can help mitigate those risks while considering patient and caregiver and also clinician or site staff participation burden, ease of implementation, and then also considering what the best sources of evidence may be. I would also add, although it's kind of a separate talk, measurement diversity can include [01:46:00] qualitative research. Reenie discussed that in the first session as well. That can also include exit interviews or exit surveys.

Dr. Anna Choe mentioned this in the first session as well, which can be really helpful for evaluating meaningful change from the patient or caregiver perspective in these trials. So to illustrate this point of measurement diversity, I do have an example for you, and Dr. Yan Wang will also discuss this example in her portion of the presentation. Here [01:46:30] we have an approval example for a disease that impacts adults with late-onset Pompe disease. It was a 52-week randomized double-blind active controlled trial. And in the measurement that I've listed here, you can see illustrated the idea of measurement diversity. There are performance-based outcomes. There is a surrogate endpoint which is sitting a forced vital capacity predicted. There are clinician-reported outcomes. There are patient-reported [01:47:00] outcomes as well. And all of these were used in constellation in the endpoint strategy here.

I want to make a couple of quick points. One, the six-minute walk test, it has labeled in more than 30 product approvals. It doesn't mean that the FDA requires it, mind you. It's an inexpensive assessment to implement. It's familiar to many site staff, it's familiar to FDA, but it isn't necessarily sensitive to detect change for every indication in every clinical trial design. And this is in part [01:47:30] why test-driving measurement is so important. I'd also like to note that this program, which was a successful product approval, also used the PROMs, patient-reported outcome measurement. And from my perspective, these are well-developed patient-reported outcomes, but the PROMs actually did not detect change in this rare disease clinical trial. And I have to say in other rare disease product approvals, I also haven't seen the PROMs detect change. It suggests to me at least that [01:48:00] patients who are living with these complex rare diseases may still be dealing with very serious impacts even when the treatment does have an effect. And it highlights for me at least that rare disease patients, especially PMD patients where the PMD is so complex, this may require a more specific measurement approach.

And so to the point I've alluded to about test-driving measurement, you could also call this pilot testing. You could call [01:48:30] this observational research. One audience member once said, "Oh, it's like phase zero of clinical trials." This is a way to evaluate and understand your measurement approach before using

it in a trial. And so this allows you all sorts of insights, how to better implement things, how to help patients learn to do the assessments more efficiently or effectively. I won't get into this in too much detail here, but it can really help reduce [01:49:00] trial failure that's due to measurement specifically. Reenie McCarthy in the earlier session had great examples of how their company learned from evaluating measurements across iterative studies and brought that forward into future study designs as well. And so my next point is just a public service announcement. All of our product approvals are public. I encourage you to read them and learn from the measurement approaches that are used. I thank you for [01:49:30] your time and attention and I'll turn this talk over to my wonderful colleague, Dr. Yan Wang.

Dr. Yan Wang: Thank you, Dr. Knoble. Today I will talk about global tests in rare disease trials from a statistical perspective. Next slide. As presented by the previous speakers, there are many challenges in drug development for PND. PND [01:50:00] are rare, complex, and multi-systemic diseases. Patients with PND have heterogeneous clinical manifestations. As a result, it may be difficult to fund a single measure to support a primary endpoint and multiple endpoint or individualized endpoint may be needed to evaluate efficacy in a trial. In addition, traditional trial designs and testing methods often have low [01:50:30] statistical power to detect a treatment effect when the sample size is small and the test product has a small or moderate treatment effect.

In my talk, I will use examples and simulations to illustrate when global tests and novel trial designs may provide a potential solution to adjust some of these challenges. Next slide. [01:51:00] Global tests are used to test multiple endpoints. The two main goals of using global tests are to increase statistical power for detecting a treatment effect and to provide a broad efficacy evaluation of treatment effects for rare diseases with heterogeneous clinical manifestations. As covered in the review paper by [inaudible 01:51:25], various global tests have been studied in the past decades. [01:51:30] The three real non-global tests Rank-Sum, Ordinary Least Squares and the Generalized Least Squares were published by O'Brien in 1984. Next slide.

This is a trial example showing when global tests may be useful. This also the example mentioned by Dr. Knoble before. This result are from an added control superiority trial in patients [01:52:00] with late-onset Pompe disease. The primary endpoint is changed from baseline in six-minutes walk test at 12 months, and the secondary endpoint is changed from baseline in FVC percent. The result of both endpoints numerically favor the test product, but the treatment difference is not statistically significant for the primary endpoint. The P-value is above 0.1. On the other hand, [01:52:30] the nominal P-value for the secondary endpoint is slightly less than 0.05. As post-hoc analysis, we conducted three global tests for the two endpoints. They are O'Brien Rank-Sum and NS-Sum, and Test-statistical Sum. The P-value of these three tests are all less than 0.03. I wouldn't go into the technical detail of this test for the sake of time, but the detail available [01:53:00] upon request. Based on this trial result, a question may be raised for future trial, should we select both endpoints as

primary endpoints and use global test to evaluate the totality of treatment effect? Next slide.

Global tests are performed by combining information from multiple endpoints. There are two ways to combine, [01:53:30] at the patient level or at the endpoint level. Two of the global tests in the previous slides, the Rank-Sum and the NS-Sum are examples of combining data at the patient level. The O'Brien Rank-Sum is based on the sum of the ranks of the outcome of each endpoint. The NS-Sum is based on the sum of the normalized scores of the outcome of each endpoint. The [01:54:00] Test-Statistical Sum is a example of combining data at the endpoint level. It is based on the sum of the two sample Test-statistic for the treatment difference for each endpoint. Next slide.

Global test has a global null hypothesis. Drug has no effect on any of the multiple endpoints. When the P-value of a global test is less than the pre-[01:54:30] specified significance level, we reject the global null hypothesis and conclude that the drug has an effect on at least one endpoint. To illustrate the treatment effect, it should present and interpret the P-value of a global test with descriptive summary statistics for the individual endpoint. One potential issue with global tests is that having a P-value less than 0. [01:55:00] 05 in a global test may not necessarily indicate an overall benefit when these current effects are observed among the multiple endpoints. Global tests have no multiplicity issue. However, they do not provide inference on individual endpoints. So having a P-value less than 0.05 in a global test doesn't guarantee a P-value less than 0.05 [01:55:30] in any of the individual endpoints. The issues here are not unique to global tests. They're also applying to the traditional test for multi-component endpoints and composite endpoints. Next slide.

In this slide we use an example to illustrate that the traditional test for multi-component endpoints are global test. [01:56:00] This example is a randomized placebo control 12 months trial in patients with Nieman-Pick disease, NPC. The primary endpoint is changed from baseline in total score of four domain NPC severity score. This four domain are ambulation, fine motor skills, speech, and swallow. A traditional test for the four domain severity score can be viewed as a global test for [01:56:30] the individual four domain endpoint, which combines data at the patient level using the total score of the individual domain endpoint. Additionally, it is worth noting that when we use a global test for the individual domain endpoint, we don't consider the total score as an endpoint, therefore there's no need to validate the total score as an endpoint. So that's the advantage [01:57:00] of global test. Next slide. In the next few slides, we use simulations to examine the power performance of a global test in comparison to the traditional testing approaches, including the Hodgeburg method and the method testing for a single endpoint.

The clinical setting of our simulations is randomized placebo control trial that use two continuous [01:57:30] endpoints for efficacy evaluation. In the first simulation study here we are assuming the test drug has the same effect side on both endpoints. The power of five testing methods are presented in the figure

here. The top three lines represent three global tests. The blue line for the Test-Statistical Sum, the red line for the NS-Sum, and the purple line for the Rank-Sum. [01:58:00] The red line here is for Hodgeburg method and the green line is for the method testing a single endpoint. This simulation shows that the three global tests are more powerful than the traditional testing methods across sample size. For example, when we have 30 patients in each treatment arm, the power of the test statistical sum is about 15% higher compared to the Hodgeburg method, [01:58:30] and about 25% higher compared to the method of testing a single endpoint. The take-home message of this simulation is that global tests perform the best in the setting when the test one has similar effect size across multiple endpoints. Next slide.

This simulation shows that the correlation among the endpoints can impact [01:59:00] the power of global tests, which increase when the correlation of the multiple endpoints increases. As shown in this figure here, when the two endpoints have a correlation coefficient of 0.2 represented by the dash lines, the power of the global test decrease by about 8% compared to the case when two endpoints are independent. The take-home message here is that [01:59:30] multiple endpoints should be selected to represent distinct clinical manifestations. Next slide.

This simulation shows that when a drug has an effect only on one endpoint, the global tests are less powerful compared the Hodgeburg method. This result is expected since including endpoints with no treatment effect [02:00:00] we will only add noise in the global test. So the take-home messages here is that endpoints that are not expected to show an effect should not be included as primary endpoint. Next slide. Now I'm going to switch to discussing another feature of global tests. Global tests can be used to provide a broad efficacy assessment for novel trials that use different endpoint for different [02:00:30] subcell patients to accommodate patients heterogeneous clinical manifestations. Next slide.

Our proposed novel trial for PMD in those patient will have different endpoints depending on their symptoms. We consider a setting where patient can be categorized into three subpopulations. In population A, patients meet the inclusion criteria [02:01:00] for symptoms of both muscle weakness and chronical fatigue and thus have two endpoints. Six-minute walk test and fatigue score. In population B, patients meet the inclusion criteria for symptoms of muscle weakness only and thus have only one endpoint six-minute walk test. In population C, patients meet the inclusion criteria for symptoms of chronic fatigue only and thus [02:01:30] have only one endpoint, fatigue score. For contrast, one may consider two traditional trial that have the same endpoint for all patients, traditional trial one, in those patients from populations A, B and C. Traditional trial two. In those patients from population A only, those 12 have six-minute walk tests and fatigue score as the primary endpoints. Next [02:02:00] slide.

The figure in this slide shows the simulation result of power calculation based on two global tests, the NS-Sum and the Rank-Sum. The top two solid lines are for the novel trial. The middle two dash lines are for the traditional trial one, and the bottom two lines are for traditional trial two. So the power of the novel trial are higher than those of the traditional trials across [02:02:30] the sample size. For example, when the sample size 15 per treatment arm within each subpopulation, the power of the novel trial is about 10% higher compared to traditional trial one and about 15% higher compared to traditional trial. Now next slide.

Now let's move on the multi-domain response Index MDIR [02:03:00] approach. This approach is a global test approach combining data at the patient level. The MDIR approach has been proposed by increased number of trial sponsors in regularly submissions in the past few years. It combines multiple endpoints using a responder threshold for minimally clinically important difference, MCID for each endpoint. There are multiple ways to combine. [02:03:30] The table shows here one example. For the outcome of each patient, the responder index score for a given endpoint have a value of class one for improvement, zero for stabilization and minus one for deterioration. The MDIR score is the sum of the index score across all endpoints. Hypothesis testing is performed to compare the means of MDIR score between treatment [02:04:00] arms. Next slide.

The MDIR approach has two potential issues. First, it may be challenging to reach consensus on the MCID threshold. And second, from a statistical perspective, it may have lower power compared to other global tests. As shown in this simulation, the power using the MDIR approach is above 15% lower [02:04:30] in the other global tests. Next slide. In summary, global tests can be used in PMD trials enrolling patients with heterogeneous clinical manifestation. They may have higher power than the traditional testing approach when the test product has effect on each of the multiple endpoints. They can provide a broad efficacy assessment for novel trials using individualized endpoints. Thank you for your [02:05:00] time and attention.

Presentation on Clinical Outcome Assessment Example

Amel Karaa, MD, Massachusetts General Hospital & Harvard Medical School

Susan Winckler: Excellent. Dr. Wang, thank you so much for walking us through that component and to Dr. Knoble as well. I have to say that actually if the two of you had been involved in my pharmacy education... Well, I retained what I was supposed to, but I might've retained even more. So thank you so much. We'll welcome you back for the reactor panel session after we hear from our next speaker, but really appreciate your contributions. [02:05:30] So our next speaker is Dr. Amel Karaa, who is an internist and clinical geneticist by training and currently serves as the director of the Mitochondrial Disease Program at Massachusetts General Hospital. Today, Dr. Karaa is going to discuss clinical outcome assessments and their importance in primary mitochondrial diseases. Dr. Karaa, the virtual stage is yours.

Dr. Amel Karaa: Thank you Dr. Winckler. And thank you for the Reagan-Udall and the FDA for putting this [02:06:00] meeting together and for accepting all my solicitations and requests over the last year. I am going to talk to you today about clinical outcome assessment lessons that we learned from recent PMD clinical trials. And I wanted to start by showing the graph of how interesting mitochondrial medicine has become over the last decade with more and more interest in pursuing clinical trials, [02:06:30] especially after 2013 when genetic testing became more available and we were diagnosing more and more patient and confirming the primary origin of their mitochondrial disease. And so initially, many clinicians started doing clinical trials that were single site clinician initiated trial in heterogeneous patient population that were clinically diagnosed, so several of them may not have had a primary mitochondrial disease. And these primarily [02:07:00] looked at dietary supplements, and they've used multiple outcome measures that were obviously not validated and really didn't show any improvement or any statistical significance.

Even after a decade of performing these trials, there are still many dietary supplement clinical trials that are investigated or initiated, who are more focused on specific subsets of mitochondrial diseases, [02:07:30] which have over the years used several outcome measures that are neither primary or secondary, but a long list of outcome measures. And I feel like these trials have played a dual purpose. One is really for the clinicians who initiated them to better understand the disease and the symptoms that the patient were experiencing and also to show efficacy of the dietary supplement that they were using. Obviously, none of them have really shown significant improvement in any of the symptoms [02:08:00] that were targeted and none of them had reached any statistical significance.

So I think part of the issue is that we have to acknowledge as clinical trialists and as investigator that a tool that is used clinically to diagnose or measure or monitor the progression of the disease, that a clinician views as a very important tool is [02:08:30] very different from what a clinical trial is perceived as a good outcome measure for a clinical trial setting. And outcome measures are really the tools used to assess the patient's status and measure a variable, for example, fatigue, which is very prominent as an outcome measure in our PMD trials. This is also very different from what sponsors and regulators perceive as a robust endpoint that has statistical analysis to help determine the efficacy and safety of the therapy being studied. And you've heard this [02:09:00] from Dr. Knoble as well. And so I think over the years we've seen that disconnect between what clinicians want to see as a tool from what the clinical biologist are advising the sponsors to use versus what the regulators and what the sponsor view as robust endpoints and outcome measures.

And that has led over the years to the choosing of multiple outcome measures that are very variable, trying to really accommodate all the stakeholders, [02:09:30] views, and assessments. So I'm going to show you a few of the trials that have been performed so far, and I really focused on the ones that were published publicly. So there are some trials that are not going to be represented

here. So obviously, when sponsors started to become interested in mitochondrial diseases around the year 2014, the lowest hanging fruit was to really go for myopathy. And because over all the [02:10:00] studies that had been done up to that point, fatigue and muscle weakness and muscle complaint were reported as being the most bothersome problem in patients and their caregivers and the one that they really wanted to be fixed first. And so obviously, primary mitochondrial myopathy became the theme for all these clinical trials.

And several programs have been invested in the treatment of primary mitochondrial myopathy. And you have heard from Reenie this morning [02:10:30] about the very long and very complicated MMPOWER program that has been ongoing for 11 years now. So obviously, a lot of these trial have used multiple primary and secondary outcome measures that have been heavily focused on muscle symptoms and signs. And as you can see here, whether the endpoint was met or not, unfortunately, [02:11:00] almost all of them did not meet their primary endpoint by the end of the trial period. We have however seen statistical significance reached when these trial went into their open label period, and in other clinical trials that were open label or with compassionate use programs where we have seen significant patient-reported improvement. And so because the initial regulators interaction have [02:11:30] been done with the neurology division within the FDA and the EMA, the focus had really been on using the six-minute walk test because this was a recognizable outcome measure and it did assess muscle performance.

And so not everyone, but the majority of these programs have used the six-minute walk test as a primary outcome for these clinical trials. And so for [02:12:00] those who are not familiar with the six-minute walk test, it's a standardized submaximal exercise to measure the distance an individual can walk within a six-minute timeframe, which reflects the integrated cardiopulmonary and muscular function. And this was initially developed for cardiac and pulmonary diseases, but over the years has also been used in other neuromuscular diseases. And as you have heard, there have been several approvals of neuromuscular disorders drugs based on the six-minute walk [02:12:30] test. But when we look specifically within our PMD population, I tried here to plot the baseline six minute walk distance, the mean six-minute walk test distance, walked by all our PMD patients in 10 different clinical trials where this was used as a primary endpoint.

And this is obviously the cross-sectional assessment of the baseline of all of these patient. And you compare [02:13:00] that with the normal range of normal healthy individual, which is between 400 and 700 meters. And you can clearly see that our PMD patient population walks less than the normal patient population about 337 meters compared to the 400 and 700. But you also see the huge variability that we see even at baseline without any intervention. And so when you try to also compare this to what happens over time [02:13:30] to our patient with PMD, this Italian cohort, we see that even if you check 12 months after progression of the disease without any intervention, you still see

that that variability exists. There is no significant change in the six-minute walk test, but the variability persists whether at baseline or after 12 months of assessment.

And the interesting thing is when you break down these patients with PMM in [02:14:00] this cohort, Italian cohort by their genotype, which is represented here in the first bar graph for single mitochondrial DNA deletion mitochondrial DNA mutation and nuclear DNA mutation or by phenotype where you see the first bar graph here and the second graph on the right, the more isolated PMM, which is represented by ophthalmoplegia versus ophthalmoplegia with proximal muscle weakness or proximal [02:14:30] muscle weakness.

Some of these patients spontaneously walk better after 12 months. So there is a huge inherent fluctuation of the six-minute walk test. And some people argued that we shouldn't use the whole six-minute score as a measure of the fatigability and the exercise tolerance in these patient. We should do it minute by minute, which has been done by the CHOP group and the Italians, and which has shown that it is indeed the patient walk less distance as the [02:15:00] test progresses. And that might be a good way of looking at the six-minute walk test. This however, hasn't been done in any of the trials that we have pursued. And so what have we learned about the six-minute walk test? There is some validity because the baseline distance in our patient with PMD is much lower than the normal range. However, the score correlation [02:15:30] between the six-minute walk and other fatigue PROs, pain PROs, and even disease severity score with the NMDAS have been very erratic from one study to another.

The reliability of the test is also very variable. Knowing that what we see is the patient who walk the least, who are very severely affected or who walk the most, who are less severely affected have the most variability [02:16:00] on the six-minute walk test. There is also limited short-term responsiveness. It is however applicable to both adults and children as long as they have reached walking ability independently. And then from the regulatory perspective, there is no regulatory guidance specifically that qualify the six-minute walk test, as you have heard, as a mandatory outcome measure to use. It is a plausible and [02:16:30] recognized functional endpoint, but it must be embedded probably within the broader outcome strategy to satisfy regulatory standards because the focus is really the patient-centricity. And do the changes we see on the six-minute walk test really translate into real benefit? So overall, we have a lot more cons to use the six-minute walk test than we have pros. And as is, it is not in my opinion, a good outcome measure, even for primary mitochondrial myopathy.

[02:17:00] So what about the twelve-minute walk test, which has also been used in one primary mitochondrial myopathy trial? It didn't really perform better even though the drug mechanism of action for that trial really called for a longer distance to be walked. So myopathies are not the only trials that have been done for mitochondrial disease. There have been several other trials that were also done in general mitochondrial disorders, not [02:17:30] focusing on

the myopathy. They have been done in Leigh syndrome, in mitochondrial epilepsy. Obviously they have used other primary and secondary outcome measures that are very variable, but have been off the shelf. Outcome measures that haven't been validated in the mitochondrial disease community. And again, we have not seen any of them reach any statistical significance and no endpoints were met.

Having said that, there are certain specific disorders within the primary [02:18:00] mitochondrial disease group that have had drugs that have met their endpoints. And those are specific for leber hereditary optic neuropathy, which is a very specific disease mutation in one specific gene. So several mutations, one specific gene that specifically affects the eye. So when you have a very focused disease that is affecting one organ system where a [02:18:30] objective measure of that organ can be done with a very homogeneous genotype, it is possible to have a good outcome measure and to reach the primary endpoint in those trials.

So overall, for primary mitochondrial diseases, we have tried all sorts of clinical outcome assessment from biomarkers to patient, observer, clinician, and performance reported outcome measure. All of them are listed here and some of them have been specifically [02:19:00] developed and validated for primary mitochondrial disease, like the Barth syndrome symptom assessment and the PMMSA for the mitochondrial myopathy that Reenie talked about. There has also been survival used for some of these open label trials, but overall these have not really been successful so far in trying to get a drug through the finish line. Some of the most commonly used one, like the problem for fatigue and quality of life are highly subjective. They are placebo prone [02:19:30] and they don't always correlate with the functional assessment. Sorry, mobility testing have been studied less in PMD, but more and more so in natural history studies so far. And they don't always correlate with each other. And the reported results vary from one study and from one cohort to another. Biomarkers would be great if we had one that was sensitive and specific for PMD. They don't always [02:20:00] correlate with clinical improvement and we are yet to see if regulators would accept this as an outcome measure. And then imaging and physiological assessment are objective. They are more sensitive, but they remain exploratory and expensive to do. So what did we learn so far from all these outcome measure? We need to validate them in very specific disease subgroups, like very specific genotype, very specific phenotype, [02:20:30] even down to the organ level. We need them to be more sensitive to early and minimal change. Like Michelle said earlier, all of these drugs that we have been testing may not have a significant of a change over the short period of time during the trial. So we need more sensitive outcome measures. We need them to be least influenced by patient perception and external factors and we need them to be easy to do and not too specialized and very specific.

So composite [02:21:00] scoring and GST approaches might be a clue here as you have heard. And then we have started to really work on these in the community. As you have heard, multiple groups have been working on

validating them. So we are on a good track. And then finally, to end, what we've learned from the trials is that it's necessary to choose an outcome measure and validate it depending on the drug that you are using. What does your drug really [02:21:30] do? How long does it take for a change to happen? Are we treating long enough to observe a change and how do we minimize the large placebo effect? And for here I would like to call on the investigators to really be more specific in how they enroll patients and eligibility criteria.And then open-label studies may not be representative of true effects in our primary mitochondrial disease community because the open-label studies have shown significant improvement. But then when [02:22:00] this trial went to phase two and phase three, those results were diluted.

So looking to the future, we really need a unified approach to CRO selection that is international, including patient voices and perspective. It's already being done. We have a huge community of international partners that are working already on unified longitudinal prospective studies internationally. These are all the stakeholders within the community and in the US we have really tried [02:22:30] to put everything together in a big infrastructure and created a clinical trial consortium called TREAT MITO which you might hear more about later. And with that I will stop and sorry I went over by a few minutes.

Reactor Panel

Amel Karaa, MD, Massachusetts General Hospital
Naomi Knoble, PhD, Center for Drug Evaluation and Research, FDA
Yan Wang, PhD, Center for Drug Evaluation and Research, FDA
Kasey Woleben, Cure Mito Foundation
Zarazuela Zolkipli-Cunningham, MBChB, MRCP, Children's Hospital of Philadelphia & University of Pennsylvania Perelman School of Medicine

Susan Winckler: Dr. Karaa, thank you so much. That was so helpful as we think about the application of what Drs Knoble and Wang had shared. And so we want to reflect on those three presentations that we just heard and to help [02:23:00] us do that we are bringing two more reactor panelists to the stage. So please join me in welcoming Kasey Woleben of the Cure Mito Foundation and there's Zarazuela Zolkipli-Cunningham Children's Hospital of Philadelphia and the University of Pennsylvania Perlman School of Medicine. I will say to Zuela and Kasey, welcome and I'm going to open it up to you. I personally have many questions [02:23:30] for what we've heard from Drs Wang and Knoble and Karaa, but your questions are going to be so much more important. Zuela, you unmuted first so you get to fire away and then we'll turn to Kasey.

Dr. Zarazuela Zolkipli-Cunningham: Okay. Thank you very much first of all to the Reagan-Udall to Amel for organizing this terrific talk and three superb speakers today. Really that was terrific. I personally am bursting with questions, so I'm going to fire ahead. My first question, if that's okay Susan, is for Dr. Knoble. [02:24:00] Dr. Knoble, you've mentioned in your talk, which was so on point, that using scales that are existing such as the PROMIS scales, which I consider generic scales, while a very

helpful resource may not be specific for disease, that is under study so our case for mitochondrial disease, which is so different to many other disorders, therefore using generic scales that are already out there may not always be the most sensitive way of capturing [02:24:30] the data.

So what do we need to do? As you stated, either we design our own or we use other existing tools. Designing our own takes time requires infrastructure. I do this in my own research, but using existing tools is really a very smart way of doing it. So before we actually decide which way we go though, you had mentioned so articulately the need to select the best outcome measures and Dr. Karaa mentioned this in her talk two.

So now my first question to you Dr. Knoble, is in [02:25:00] the process of demonstrating why we selected those particular outcome measures, there's been a mention of a need for qualitative interviews. Do we need to publish that material prior to conducting the trial or submitting to the FDA? Bearing in mind everybody has initiative to take trials off the ground very quickly. My second question is because we're aware of data that is out there in related disorders, for example, Friedreich's ataxia [02:25:30] where my colleagues shared, they have two decades worth of data, natural history that's been quantifiable in showing decline in the nine-hole peg test assessment of dexterity. However, there it was shared that patients don't use pegboards in their daily life. Maybe that's not the best outcome measure. How would we go about demonstrating clinical meaningfulness for this when we understand that the nine-hole peg test, at least in mitochondrial disease, reflects ability to do buttons and put in contact lenses and earrings and so forth? [02:26:00] So those are my two questions. Thank you.

Dr. Naomi Knoble: Thank you for your wonderful questions. And forgive me if I didn't encode them all immediately. So if I ask you for some clarification, please don't take offense. To the first question around identifying what it is that you want to measure. Dr. Zolkipli-Cunningham, one of your publications is one I routinely cite in presentations where you did a survey [02:26:30] with both children and adults and in your tables in the publication. But my favorite was the supplementary tables where you broke it down and showed that what children or caregivers of children and adults with PMD were reporting is as their key symptoms differed in the way they may have arranged the top four, but in fact the top four symptoms were the same.

Well first of all, it was great work. Second of all, it's a great illustration [02:27:00] of helping to design trials and as that really important starting point for identifying what matters, it's great when you can find things in common across kids and adults and it's tremendous when you can find a way to actually use the same measurement that reaches both populations as well. It's ideal. It's very tricky sometimes though. You had asked a question in regards to the qualitative research. Qualitative research, [02:27:30] like interviews, we'll call it concept elicitation research, sometimes we'll call it cognitive debriefing or cognitive interviewing. You could also use probably social media research.

There's a lot of ways to go about qualitative research methods. We have a guidance about it. It's guidance two in our patient focused drug development guidance series.

You can mine that for lots of, I think, insights. One to my point in the talk about getting specific, what is the specific aspect [02:28:00] of... Let's just use fatigue. What's the specific aspect of fatigue that people are struggling with? I heard, I think, in one of the talks from the first session, it's the recovery time after exertion and maybe that's the impact, that patients can exert but then have this really negatively impactful recovery time and maybe we can shorten that recovery time or eliminate it altogether [02:28:30] for things that don't even seem necessarily exertional to the average person.

So I think your question was also in regards to the sequence of publishing. To the best of my knowledge, FDA doesn't say, "Oh, you have to publish this first, then this." Things can be done iteratively and often are certainly in the IND, the investigational new drug, phase of research of clinical research and that's completely [02:29:00] acceptable. It's great to publish these things. And to your point in using, I think you used the Friedrich's ataxia patient community, being able to mine existing evidence, whether it's quantitative or qualitative, it really can be helpful, again, for just that iterative process of how can we glean more insights? You had a second part of your question. I haven't retained it.

Dr. Zarazuela Zolkipli-Cunningham: Oh, that's okay. I was just wondering it's the same correlation to the first question [02:29:30] all about showing the insights from the patients and clinical meaning. So if we are going to say, for example, I've been following patients since 2016 and do have the nine-hole peg test data showing very slow decline over time, what would we need to use this as a, let's just say there was a need for measuring dexterity in a future trial. Would we need to also conduct qualitative interviews? Would we [02:30:00] need to show clinical meaning in some way and publish it prior to submitting this as an outcome measure?

Dr. Naomi Knoble: Thanks for reiterating again. Again, the sequence of publishing and demonstrating clinical meaningfulness. That sequence isn't always there as you even noted, right? Companies start clinical trials and that's okay. My colleague Dr. Anna Cho mentioned this in her talk and I often stand on this soapbox as well. You can [02:30:30] use exit interviews or exit surveys as a way of integrating qualitative evidence around clinically meaningful change into a clinical trial. So if you've had a randomized blinded period of your study, at the end of that, being able to interview caregivers with patients in an interview or in a survey just to find out what impacts they did experience. And if the nine-hole peg probably if it's not going to show change in a [02:31:00] short duration, but it's used as exploratory in the study for people to be able to be interviewed and ask, "Did anything change for your fine motor skills? Were there any improvements or worsening? What were you able to do if there were improvements? What weren't you able to do if there was worsening for you?"

And being able to understand those kinds of insights. Patients are the number one stakeholder in everything we're doing here and [02:31:30] from the end of a clinical trial. They are the number one source of insight in what it's like to take this product and what it did for them or didn't do. And so I think it's a great tool, a great qualitative research tool to use. And we do talk about that also in our patient-focused drug development second guidance too.

Dr. Zarazuela Zolkipli-Cunningham: Thank you.

Susan Winckler: That's great. What a great tee up of question and Dr. Knoble, I think you tracked through what I counted as a four-part question, so that was great to [02:32:00] further explore that. I want to turn, Kasey you listened as well. We want to have you jump into this reactor panel also.

Kasey Woleben: Yeah, thank you so much Dr. Winckler for inviting me for today's panel and thank you for the presenters. They were awesome presentations. I did like the daily diary. I would like to do a video daily diary. [02:32:30] I think that for children, I'm going to be on the pediatric side just because my group's primary focus is on Leigh syndrome, which is the most common pediatric mitochondrial disease and it is neurodegenerative. So to me, I'm hearing all of these outcome measures, like the six-minute walk test thinking our kids are losing their abilities. Some of them haven't even walked. So it's just disheartening when I [02:33:00] hear the six-minute walk test as a primary endpoint. My question is though, for rare and treated disorders, the FDA allows approval based on biomarkers. There are no validated biomarkers for mito disorders. What would the FDA anticipate a biomarker in PMD to look like? And I'll just throw it out there.

Susan Winckler: Yeah, I just say any thoughts on that? I mean, Dr. Karaa, you've certainly thought about this space quite [02:33:30] a bit. Do you want to jump in and offer some thoughts?

Dr. Amel Karaa: Yeah, I think biomarkers are really great when they are really specific and really sensitive and we can correlate them to disease severity, progression, improvement and how the patient feel. And so if a biomarker fits all of these, then of course great when I'm sure the FDA would agree to use that as an endpoint, but we're not there yet. [02:34:00] And I think for mitochondrial diseases it's very hard because we know how complex and heterogeneous these disorders are. So coming up with a unified biomarker that would fit all is going to be really hard.

So as Dr. Knoble said, I don't think there is a perfect outcome measure out there that we can rely on or that we can invent or validate. This is a group of disorders that will require a more surgical approach, [02:34:30] a more targeted approach with even going down to subgroups with subgenotypes focusing on one organ and that's how we can make it work. And then once we have that one first drug approved, then we can start expanding and looking into more organs and into more systems and into more phenotypes and genotypes. It's been really difficult, but it's not discouraging because [02:35:00] everyone, all the

stakeholders have really been hanging in there and have been really doing great work. We're collaborating more, we're doing more data analysis like Marni said earlier. It's really important to look at the data and aggregate all the data that we've gathered so far to be able to analyze and move forward with making informed decisions based on data.

Susan Winckler: Yeah, I'm struck both by Kasey's question and your response Dr. Karaa, that this is space where [02:35:30] having that ability to say, "What can we learn in some of these discrete areas and then apply more broadly," is helpful. But then recognizing as noted that there are very discrete areas and we have to explore those specifically. Dr. Knoble, you unmuted, so I want to turn to you as well.

Dr. Naomi Knoble: Thank you. Yeah, first I just want to say Dr. Karaa, your optimism is infectious. And I think if we see nothing from the entire PMD community, [02:36:00] it is both perseverance and optimism. Kasey, your question about biomarkers is great and to Dr. Kara's point, although it's tricky, it doesn't mean it's impossible and, obviously, there's perseverance enough in this community to, I think, make it possible. Biomarkers come in different types and it's very tricky to find a complete and total biomarker that gives us the whole of the disease's pathophysiology. Often what we find are these partial biomarkers, and [02:36:30] in that instance you have this one piece of evidence that say, "Well, we moved a little piece of this disease." But it might not tell you everything that also really matters to patients and families.

Like an example that comes to mind is joint X-rays for maybe some type of arthritis will help you understand the progression of how that joint is deteriorating, but it won't tell you about pain and swelling and that's so important for your functioning, right? So these [02:37:00] are just examples, but I think curative treatments are obviously the goal for everybody. And when treatments are going to just impact a specific symptom, we're going to need this kind of measurement diversity. And so I imagine future of clinical trials for PMD that will hopefully include biomarkers, but will likely also continue to include clinical outcome assessments. That's my best guess.

Susan Winckler: Yeah, excellent. Thanks so much. We can turn back to you [02:37:30] Dr. Zolkipli-Cunningham and I should say you can also unmute and jump in as we're having the conversation. So go ahead and Kasey be ready.

Kasey Woleben: Well, I do have a comment. So our Leigh syndrome is such a severe disease and our children are dying and it's just how can we move things forward quickly? For a degenerative disease, just making [02:38:00] it stop the progression, that's a win for us. Our children may not be able to walk, talk, or eat by mouth again, but to have the disease just stalled out, that is a huge success for us. So, I guess, I don't know. I know that we're moving things forward and everybody's working together quickly. It's just, as a parent, I just [02:38:30] urge that our children need this so...

Susan Winckler: Kasey, thank you for investing your time and energy in this engagement overall and I was struck when you noted the importance of a video diary, right? It's far easier in capturing that, particularly [02:39:00] in a population where they may not be able to provide a response themselves, but you could engage the parents or the caregivers in doing that. So appreciate that. Zuela, I'll turn to you.

Dr. Zarazuela Zolkipli-Cunningham: Yeah, thank you. I think I have a question for Dr. Wang, but I think on Kasey's note, which is so heartfelt and we all need to listen to our patients, I wouldn't mind sharing, if I may, everyone, it's not so much to do with outcome measures. But [02:39:30] touching on what Kasey stated, I very gratefully have Elizabeth and Jeff Reynolds from The Champ Foundation have shared with me their sentiments on patient enrollment to clinical trials of the pediatric population where it's very hard because most of the trials that Dr. Karaa showed were primarily in adults, particularly when it's primary mitochondrial myopathy. I'm just going to read what they've shared with me because it's stuck with me. " [02:40:00] We live in constant fear of the unknown. Families engaging in research, live experience, and understanding of an exceedingly complex medical condition make us the most qualified decision makers. We hope that future PMD clinical trials will be open to pediatric patients. Families are capable of making the decision to enroll or not and deserve the right to make that choice." It's not necessarily a question, but it's a statement [02:40:30] I wanted to share.

Susan Winckler: Very powerful. Thank you.

Dr. Zarazuela Zolkipli-Cunningham: Great.

Susan Winckler: Did you want to ask Dr. Wang a question? You may.

Dr. Zarazuela Zolkipli-Cunningham: Absolutely. So now I'd like to get into the itty-gritty Dr. Wang. So I love statistics, I am by no means a statistician. I will share that I loved hearing your talk on global tests. So I have some comments, if I may. The first thing I did want to say is as a non-statistician, to me, global tests [02:41:00] feels different to a composite score, right? So I have the [inaudible 02:41:04] composite score there. You decide before the trial and on submitting to the IND that you are going to go ahead and conduct all the assessments and then come up with a single unifying composite score. It feels like with a global test, it's more post-hoc. So sort of like in a non-clinical trial it would be a meta-analysis, right? You're looking at the data you have and then you're combining it to increase the power. [02:41:30] That sounds terrific. Absolutely.

The only thing I will say is on one of your slides, it did state an example of where the N was 20. So, again, this raises the concern that the cohort in PMD, the available cohort in PMD might be too small for such an analysis. That's one point to make. The other point to make is if you go about and state that you will use the global test approach, does that mean that we can't then consider [02:42:00] the individual endpoints? Turning it over to you.

Dr. Yan Wang: Thank you very much for the questions. I think the first question you are talking about the example we show post-hoc. That's true because I think that study was done a while ago, but we do have new trial using global tests as the predefined primary analysis. [02:42:30] Unfortunately, there's no published data we can share for now at this moment. So global tests can be used as predefined primary analysis, not just using as post-hoc analysis.

And second to your question, no, I think global test, yes, in the layman's sense you can think about it's almost like a meta-analysis, right? The [02:43:00] meta-analysis here is for different endpoints, but that's why it feels useful in PMD cases because it's unlikely, as many speakers spoke about, very difficult to identify one single endpoint applying to all the patients. So this global test can apply in [inaudible 02:43:27] trial design in the sense that allow [02:43:30] patient have individualized endpoints as the example we show in one of the slides. And let's do that. The global test, combine the information from those individual endpoints together, make a statistical inference. So I think that makes sense when I think about as layman. What's your third question?

Dr. Zarazuela Zolkipli-Cunningham: Yeah, actually [02:44:00] on that point, and my third question, and I'd love Dr. Karaa to weigh in here. So the Nobel trial, the slide that you showed up of the Nobel trial design potentially in a mitral disease patient, my only comment there from knowing our patient's evaluations is where you had stated, well, perhaps the person with muscle weakness of that assessment and the person with fatigue have that assessment. Did I misunderstand? I think it is important for me to point out that the majority of [02:44:30] our patients have a combination of symptoms. They don't have symptoms that are independent to each other, which makes that quite tricky for them to choose which one because on Monday it might be one, and on Friday it might be the other one.

So I think it's a little dangerous to say maybe choose just one, but certainly two or three would be a better approach and be more reflective of the fact that mito disease patients have an average symptom of 16 symptoms and it's never just going to be one. So I think that would be just be my only caution [02:45:00] about your suggestion of perhaps choosing one and then taking the analysis forward. But we'd love to hear from Dr. Karaa too. Thank you.

Dr. Amel Karaa: No, I agree with you as well completely. It's very hard for patients even to understand why we have chosen these outcome measures. We've actually done video exit interviews in many of these trials. We have even done video interviews to ask patients what they thought about [02:45:30] these outcome measures and whether they really captured what they were feeling and I have been very surprised at what I've heard from patient, including with a six-minute walk test. A lot of these PMM trials, we enroll patient who have ptosis and Ophthalmoplegia VI, which is a myopathy, and they're telling me, "I don't care about how much I walk. My disease doesn't impair my walking, it impairs my vision and you're not asking me anything about my vision." [02:46:00] We've had patient ask about the most bothersome symptom like, "Every day I have a different most bothersome symptom. I can't really choose one at the beginning

of the trial and stick to it because my disease fluctuates day to day and so this might not be the one that I feel most affecting me on the next visit."

And they argue with you at the next visit because they're like, "I want to change my most bothersome symptom, can I change it?" And so I agree with you, there [02:46:30] isn't really one outcome measure that we should be using and that's why I said is moving forward composite scoring and GST methods most appropriate for our patient? I understand that this is a really complicated place to trying to figure out what these outcome measures should look like. But I think a lot of people bear responsibility to work together. There are the key opinion leaders who [02:47:00] usually advise the sponsors as to what the trials should look like and what outcome measure we should develop. They all have different expertise, they all are seeing patient from a different lens and they all are putting their research and clinician hat on where in fact these opinion leaders really need to think about the purpose of the trial and why we're doing this trial and really think of outcome measures that are not clinically oriented, but more trial oriented and which include [02:47:30] the patient's voice and the patient patient feeling and how the patient live their day-to-day life.

So a huge responsibility is from these key opinion leaders and how they view trials and how they approach outcome measures. I think the second group that has responsibility are the sponsors. And even though I think most sponsors that are working in the rare disease community genuinely want to help patient and [02:48:00] find the drug that treats patient because they really understand how severe these diseases are, especially in the pediatric population in cases, to your point, I know for a fact a lot of them are working on pediatric trials, but these are even more complicated than adult trials. So my comment to the sponsors is that it's really important to spend time working on all these details before the trial starts.

And I [02:48:30] understand that it's very costly and there are pressures coming from different places to move forward quickly, but the money and the time you will spend upfront really ironing out all these details and what's the most appropriate outcome measure and should we do cognitive debriefing and qualitative interviews to really hone down that perfect, not perfect, but the closest to the best that we can get will really pay off at the end as opposed to starting a trial, not [02:49:00] being sure really where we're at, and then having to change things along and not reaching your end point at the end.

And then regulators are the third group of responsibility, right? So I know that the FDA has really tried to respond to all the challenges of rare diseases, but I think primary mitochondrial disease within that rare disease group really stands alone because I don't know if any [02:49:30] other disease that is as complex as heterogeneous and trying to fit these PMD programs into this mold doesn't really work. So I feel like regulators need to be more open and more thinking outside the box when it comes to trials for PMD because we can't really do it the same way where we do it for other trials. And so if all these stakeholders and the three groups can agree at least to work [02:50:00] on these challenges

together, we can get there. We can find a way to find the best outcome measures for our patients.

Dr. Zarazuela Zolkipli-Cunningham: And I will say there on that note too, Amel, sorry. The really important thing is we need to be prepared and ready because in other disorders, they're struggling with understanding what outcomes are from combined therapies. We're not there yet, but hopefully we will be someday. And we really need to be able to work together and be prepared industry [02:50:30] partners, stakeholders, families, to be able to actually conduct the studies, how do we do it to understanding what the outcome measures, what the outcomes are when we have one, two, three, four drugs that are approved.

Susan Winckler: Yeah.

Dr. Amel Karaa: Yeah. so to that note, and sorry Susan, if I may, we have listened to industry, right? In partnership with the UMDF and as part of the Treat MITO Clinical Trial Consortium we have started performing these outcome measures at the conference, at the UMDF [02:51:00] conference through the clinical research pavilion. So every year now at the UMDF Annual Conference, we have a huge space put together for patient to come and do all these assessments. And Zuela, you have been instrumental in getting your MM coast studies done there. And so this is a way for us to really to respond to these needs and to open the door to patient caregivers, all comers [02:51:30] to come try these tests and tell us what they think about them and whether they are capturing their ailments every day. And I hope that we will continue to do this every year and we will extend it. We have already multiple industry sponsored studies prepared for this year, and we are planning on doing more over the years. So I think this is one way of really responding to this challenge.

Dr. Zarazuela Zolkipli-Cunningham: Yeah, that's right. And I think Dr. Knoble, just so you know, we're test-driving, I think I like your term. And we're test-driving it so Phil Yeske-

Dr. Naomi Knoble: Test-driving, that's a great example.

Dr. Zarazuela Zolkipli-Cunningham: Exactly, so Phil [02:52:00] Yeske's been supporting our group. I bring a CHOP physical therapy team. We've actually been testing out new digital technologies as was discussed in the first panel today. So really doing a test drive, collecting data in real time at a patient conference. We collect data of 70 individuals for the last two years, and we're going to do it again this year. So test-driving truly is engaging community effort for sure and it's being done in our community.

Susan Winckler: [02:52:30] Yeah, that's great. So I'm going to jump in here and note that we are just a little bit over time by the clock, but Kasey wanted to give you, if you wanted make one final observation that will help us close out the morning and then we'll take our break and return this afternoon. So Kasey, did you want to offer one additional thought here? I'll give you that stage.

Kasey Woleben: I'll just kind of repeat what Dr. Karaa said, FDA, think outside the box. I know mitochondrial diseases [02:53:00] are complex and our children deserve clinical trials to help them. So thank you so much for this opportunity.

Susan Winckler: Yeah, and I think Kasey, as I hear you, I connect what each of our speakers has said this morning, and it was echoed by Drs Knoble, Wang and Karaa about being intentional and having the conversations between the researchers, the academia, [02:53:30] those key opinion leaders, the sponsors, the patient community, and then having those conversations with the regulators so that we're all keeping pace with the developments. With that, I'm going to close us out and we are going to restart in 25 minutes. So on the East Coast, that's 1:30, it's adjusted for wherever you might be in the world, but know that in 25 minutes we are going to restart and kick off for the afternoon. So thank you so much, Drs [02:54:00] Wang, Knoble and Karaa as well as Zolkipli-Cunningham and Woleben. We really appreciate your contributions.

Dr. Amel Karaa: Thank you.