U.S. FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

AND

JOHNS HOPKINS UNIVERSITY CERSI WORKSHOP

Addressing Challenges in the Design and Analysis of Rare Disease Clinical Trials: Considerations and Tools

Day 1

Tuesday, May 2, 2023

9:00 a.m. to 12:02 p.m.

Meeting Roster

John Concato, MD, MS, MPH

Associate Director for Real-World Evidence Analytics, Office of Medical Policy, Center for Drug Evaluation and Research (CDER), U.S. Food and Drug Administration (FDA)

Sorin Fedeles, PhD, MBA, MS

Executive Director, Polycystic Kidney Disease Outcomes Consortium, Critical Path Institute

Kerry Jo Lee, MD

Associate Director for Rare Diseases, Rare Diseases Team, Division of Rare Diseases and Medical Genetics (DRDMG), Office of Rare Diseases, Pediatrics, Urologic and Reproductive Medicine (ORPURM), Office of New Drugs (OND), CDER, FDA

Christine Nguyen, MD

Deputy Director, ORPURM, OND, CDER, FDA

Caitlin Nichols, PhD

Research Director, AllStripes Research

Aliza Thompson, MD, MS

Deputy Director, Division of Cardiology and Nephrology, Office of Cardiology, Hematology, Endocrinology, and Nephrology, OND, CDER, FDA

Vanessa Vogel-Farley, BA, BS

Senior Director, Research & Data Analytics, Global Genes and Principal Investigator, Rare-X Data Collection Platform

Ramona Walls, PhD

Executive Director of Data Science, Critical Path Institute

Scott Winiecki, MD

Team Lead, Rare Diseases Team, DRDMG, ORPURM, OND, CDER, FDA

C O N T E N T S

C O N T E N T S (continued)

P R O C E E D I N G S

(9:00 a.m.)

Welcome

DR. LEE: Hello. My name is Dr. Kerry Jo Lee, and I am the Associate Director for Rare Diseases in the Office of New Drugs, Center for Drug Evaluation and Research, or CDER, and lead of the Rare Diseases Team, which manages CDER's Accelerating Rare disease Cures or ARC program.

I am very happy to welcome you to this FDA CDER and Johns Hopkins University Center of Excellence in Regulatory Science and Innovation Workshop, entitled Addressing Challenges in the Design and Analysis of Rare Disease Clinical Trials: Considerations and Tools.

This workshop is one of several events under the umbrella of CDER's ARC program, which in its first year is focusing on engagement with stakeholders, both to better understand their challenges in designing and conducting clinical trials in rare diseases, as well as to inform and share FDA's current thinking on regulatory

considerations regarding these trials.

I am personally very excited about the program we have put together for you over the next few days. There remains a tremendous unmet need for approved therapies for rare diseases that affect between 25 and 30 million Americans. That means about 1 in 10 Americans have a rare disease. And while collectively this is not a small number of people, when it comes to developing therapies in very small populations, there remain a number of common challenges that's imperative that we remain thoughtful about the collection, use, and analysis of the data that we receive because in small populations, every patient's experience is critical to both informing trial design, as well as demonstrating a potential therapy's effectiveness.

This workshop will share experiences, best practices, and the regulatory perspective on how to collect high-quality and fit-for-purpose data for rare disease clinical trials; the use of data sources to inform rare disease drug development; and design and analysis methodologies for use in

rare disease clinical trials. My hope is that you will take away something from today's program that will better help you to advance your own work in developing safe and effective therapies for rare disease patient populations.

Without further ado, I am going to turn this over to the first session moderated by Dr. Scott Winiecki. Dr. Winiecki is currently a team lead on the Rare Diseases Team. He is an experienced pediatrician who trained at the Children's Hospital of Philadelphia. He has been with the FDA since 2011, with experience both as a reviewer in the Center for Biologics Evaluation and Research, as well as CDER's Professional Affairs and Stakeholders Engagement staff, where he led the Safe Use Initiative to reduce preventable harm for medications through extramural research.

Dr. Winiecki, I turn it over to you.

Session 1 - Scott Winiecki

DR. WINIECKI: Thank you, Dr. Lee.

Our first session is about how to collect high-quality and fit-for-purpose data. We live in

an age where many rare disease advocacy groups have started to collect data via natural history studies or registries, and without question, this data is crucially important in the context of rare disease drug development. However, this data needs to be collected and organized in a way so that it can be most useful for understanding rare diseases, for structuring clinical trials, and for regulatory submission.

This is what our first session is all about. We're going to have three talks today in this session, all covering data collection and data organization. I'd like to remind everybody that during the panel session, we will be answering questions, some that were submitted when you registered, and others, if you think of them as the talks are going on today, please enter them in the Q&A box, and we will cover as many topics as time allows in the panel session.

Now, I'd like to introduce our first speaker, Dr. John Concato. He is the Associate Director for Real-World Evidence Analytics in the

Office of Medical Policy. Dr. Concato joined FDA after a 27-year career at the Yale University School of Medicine, as well as the U.S. Department of Veterans Affairs. At FDA, his responsibilities include a focus on FDA's real-world evidence program and include looking at internal agency processes; external stakeholder interactions; demonstration products; as well as guidance development. He also serves as the chair of CDER's Real-World Evidence Subcommittee.

Today he's going to speak on Regulatory Perspectives on Real-World Data, and his talk will highlight several FDA guidances, which reflect FDA's current thinking on real-world data and real-world evidence.

Dr. Concato?

Presentation - John Concato

DR. CONCATO: Thank you, Scott and, thank you Kerry Jo, and thanks for inviting me to this program. I'll be talking, as mentioned, on regulatory perspectives regarding real--world data. Next, please. The views and opinions are my own

and should not be attributed to FDA's official policy. I do not have any conflicts of interest to report, and if I mention a commercial product, it's not an actual or implied endorsement. Next.

Just to give you a sense of the flow of this presentation, I'll first start with a bit on historical context, leading to the current use of the terms "real-world data" and "real-world evidence." I'll spend most of my time describing the main components of FDA's real-world evidence program, emphasizing guidance development, and then I'll close with a few slides on challenges and potential contributions of using real-world data and real-world evidence in general, as well as for rare disease. Next, please.

Just to start, these definitions of real-world data and real-world evidence come from our 2018 framework. On the left, we see that real-world data are data related to patient health status or delivery of healthcare, routinely collected from a variety of sources. So for a very simple definition, you think of electronic health

records, medical claims data, data from registries, et cetera.

On the right, real-world evidence is evidence derived from the analysis of real-world data; again, a simple definition. Importantly, in the lower-right corner of the slide, often overlooked, various study designs can generate real-world evidence, including randomized trials in certain circumstances, but certainly externally-controlled trials and observational studies perhaps come to mind first.

Here's a bit of historical context outside of drug development per se. Let's think of the term "Big Data." That first appeared in the computer science literature, actually, during the 1990s and initially referred to data just too large to be stored in, then, conventional storage systems. If we fast-forward -- it's already more than a decade ago but -- into the 21st century, big data represents, quote, "shorthand for advancing trends in technology that open the door to a new approach to understanding the world and making

decisions," close quote.

So one perspective is that as modern technology has advanced, we have increased quantity and forms of available data, as well as, importantly, the speed to merge and manipulate the data. But we should remember that integration and analysis of large-scale data has always been integral to epidemiology and drug development science. Next, please.

Here we encountered the 21st Century Cures Act of 2016, where FDA was mandated by Congress to establish a program to evaluate the potential use of real-world evidence to support a new indication for a drug already approved or to satisfy post-approval study requirements. That same framework I mentioned was issued in December of 2018, and we followed up with draft guidance for industry in late 2021 and thereafter.

I think it's important to emphasize that our standard for substantial evidence remains unchanged; that is whether evidence comes from a trial, a traditional randomized trial, or a

so-called real-world evidence study. And we don't have time today, but commitments were met under the Prescription Drug User Fee Act VI, and we're on our way with PDUFA VII. Next, please.

That 21st Century Cures Act is perhaps an inflection point regarding the use of the term "real-world evidence." It actually is a nonspecific modifier. Real-world data and real-world evidence appeared in the medical literature as of the 1970s or earlier, but in various unrelated contexts. The contemporary usage, however, now has specific regulatory implications. So one way to look at the situation is older epidemiologic terms were just fine, but the emergence of big data that I described, as well as the enactment of the 21st Century Cures Act, has led to where we are now, that is actually sometimes confusing use of different taxonomies or descriptions of study design.

The main point I want to make right now -- and I'll circle back to this later -- is when you hear RWE study, that's not synonymous with

observational study. You really need to know additional details to understand what study design is being used or described. Next, please.

So here's where I pivot to FDA's real-world evidence program after that general background. I want to emphasize this applies to the Center for Drug Evaluation and Research and Biologics Evaluation and Research, as well as the Oncology Center of Excellence, for drugs and biologics, that is, across the board. We get along quite well, and we collaborate with our Center for Devices and Radiological Health and other centers, but they have their own regulations, and therefore, they have their own guidance on real-world evidence. The drug and biologic programs can be described informally in four categories: internal agency processes; external stakeholder engagement; research AKA "demonstration" projects; and guidance development, and the next series of slides will walk through these four categories. Next, please.

Actually, the first and second categories are on one slide. I just want to highlight the

Real-World Evidence Subcommittee and its role in supporting internal activities. The membership of that subcommittee is FDA staff, including leadership for multiple CDER and CBER offices. It provides oversight of policy development on real-world evidence, including guidances that I'll be describing. It offers resources in leadership to review divisions, among other activities.

In terms of external engagement, the committee provides feedback on early-stage proposals, not drug development per se, but rather novel ideas for new data collection, et cetera, cross-cutting ideas from sponsors or vendors. It also discusses initiatives presented to the subcommittee for consideration, and then there are additional activities such as holding FDA- or Center-level public meetings, or conducting small business and industry webinars, or speaking engagements such as this morning. Next slide, please.

If we turn just a slide or two on demonstration projects, here's where FDA is

investing in the future by funding projects that focus on data, study design, or tools, including via CERSI mechanism and other funding award mechanisms. I have six examples here listed. I think I'll just read across, left to right, for the first row, for the interest of time.

In terms of improving the quality or use of real-world data, the OneSource project with the University of California San Francisco is a project to improve the quality of EHR data. Why wouldn't clinicians want research-grade data at the bedside? That is one way to look at that project.

In the middle column, study design, the acronym RCT-DUPLICATE was a study of observational data. Actually EHR and mainly claims data was an observational cohort design to see if the results of randomized trials could be emulated. For those who are in the field, you might know that last week, in JAMA, the Journal of the American Medical Association, the Main Results manuscript was published, and I encourage folks to read that article if they're interested.

The right-hand column and the first bullet point under the tools category, we see evaluation of confounded treatment effects. If a study isn't a randomized trial, we worry that - the technical term is called, "confounding," where the result might be biased. This project funded a group at the University of North Carolina to look at how we have a better sense of how to use an approach to assess how much that confounding might impact the results. Next slide, please.

Here, I will go directly to guidances and spend about 8 or 10 slides discussing this topic. I will say upfront, these four screenshots are four of our main guidances for real-world data and real-world evidence. It should be apparent, as I walk through these slides, that we used a modular approach, one might call it, or a reductionist approach. Rather than try to write one single uber guidance that would be very long and very complicated, this is sort of one-stop shopping in the sense of when you want to know about data.

Let's look at the left-hand side of the

slide. Assessing electronic health records or medical claims has its own guidance, and below that, assessing registries. FDA's current thinking is reflected in those two guidances in terms of data sources.

On the upper right, data standards, we recognize that our data standards and our regulations anticipated clinical trial data. What do we do when we have data coming from real-world data sources? Well, this guidance helps explain that. Then on the bottom right, considerations for the use of real-world data and real-world evidence to support regulatory decision-making. Our regulations, again, anticipated clinical trials. What do we do if the design is observational? So, next slide.

Here, I'll start walking through those four guidances one at a time. This is a screenshot of the title of our so-called EHR claims guidance. Next, please. As an overview, the focus of this guidance is on selecting data sources to appropriately address the study question with very

granular details on development and validation of definitions for exposures, covariates, and outcomes, and recommendations on data provenance during accrual, curation, and analysis, and study design is handled elsewhere. Next, please.

This is the cover page of our, quote/unquote, "registries" guidance. Next. Here's where if a stakeholder is working with registries, we describe registry fitness for use in regulatory decision making, focusing on how to collect relevant and reliable data. Very often when using registries, linkage to other sources for supplemental information, such as claims, EHRs, and digital health technologies is involved, and we have recommendations in that regard. Then finally, we have a section on FDA review of submissions that include registry data. Next, please.

The data standards is the third of four core guidances from 2021. Next. Here's where we describe processes for managing real-world data and how to conform real-world data to FDA data standards -- again, that anticipated clinical

trials, mapping the real-world data to submission standards, and considerations for data transformations. Now again, this is a technical guidance, but it applies regardless of the type of real-world data; and certainly in terms of sponsors listening to this conversation, there are teams involved that would have the requisite expertise. If patient advocacy groups are listening, it's a question of making sure that the time, effort, and trouble of collecting the data is worthwhile, so we encourage early engagement with the FDA in that regard. Next, please.

This is the fourth of the core of four, regulatory considerations guidance. Next. Here's what I already alluded to: marketing applications to support the safety and effectiveness of a drug must satisfy legal standards, even if the 21 Code of Federal Regulations part 312 involving investigational new drugs does not apply. So our so-called IND regulations in part 312 did not anticipate the era of real-world evidence, but this guidance fills in the gap.

I will mainly say that there are two classifications of non-interventional studies. One involves only the analysis of data on the use of a marketing drug in routine practice. Secondly, there are ancillary protocol-specified activities or procedures. The drug could be given in clinical care but additional lab tests, imaging studies, or questionnaires might be performed, say, in a natural history study.

FDA does not consider these types of studies to be clinical investigations but, nonetheless, protection of human subjects is critical, so sponsors must meet the applicable requirements under the FDA regulation shown at the bottom of the slide in terms of protection of human subjects and institutional review boards. Next slide, please.

I'm now going to cover a few additional guidances that came out after 2021. This guidance on externally-controlled trials was published several months ago in 2023, and the next slide shows that the content emphasizes the importance of design considerations such as finalizing a protocol

before analyzing data; specific data considerations for the external control arm, various comparability issues; specific analysis considerations, and although FDA does not recommend a particular approach, it's basically picked the right tool for the job rather than us saying that a specific approach is better than all others in all circumstances; and then considerations to support regulatory review or access to patient-level data so we could do our job in the review mode.

Just as a technical note, this guidance does not address external control data based on summary-level estimates; rather, it's patient level, and it also doesn't address supplementing a control arm in a traditional randomized trial. The last scenario sometimes goes by the name of a hybrid randomized controlled trial. Next slide.

I really want to emphasize this point. It's from the external control guidance, but it really applies pretty much across the board. I'll read or paraphrase most of the text there.

Sponsors should consult with the relevant

FDA review division early in a drug development program about whether it is reasonable to conduct an externally controlled trial, or fill in the blank, instead of a randomized-controlled trial. As part of these discussions, sponsor should provide a detailed description of the reasons why the study design is viewed as appropriate; proposed data sources, and an explanation of why they are fit for use; planned statistical analyses; and plans to address FDA's expectations for the submission of data.

This, again, is a very pivotal point to make, so we try to share this every time we get a chance to speak externally. Next slide, please.

I also want to mention a procedural guidance "Submitting Documents using Real-World Data and Real-World Evidence to FDA." Next slide. I won't say much about that guidance, other than the main point with this guidance is that you could help us to help you -- as sponsors especially -- by in your cover letter, indicating exactly what is involved with the real-world data or real-world evidence.

All too often, we see false positives where the terms are just thrown in, or false negatives, where it's saying an externally-controlled trial is submitted and real-world evidence is not used. We could always update that for classification purposes, but since we have a mandate to report to Congress, it would be more efficient for everyone to adopt a standardized approach. Next slide, please.

The next slide is a chance for me to just summarize where I've been. If we look in the left-hand column, we see that the modular approach to guidance development is such that we have two guidances on data considerations themselves; one guidance on data standards for submission of data; a uber guidance, or an overarching guidance I should say, on the applicability of regulations; and then only 1 of 3 in our design category where the externally-controlled trials guidance has been published.

Please be aware that for a trial in practice settings, non-interventional studies guidances are

in development and will be going through the clearance pipeline in the near future. And last but not least, the procedural guidance that I mentioned was published in September of 2022. Next slide, please.

Not necessarily an RWE guidance, but Digital Health Technologies for remote data acquisition and clinical investigations, this gives me a chance to mention this guidance that was also generated in December of 2021. Next, please.

Here, I'll stop with the guidances and just try to bring us back to a more overarching view of real-world evidence. This article is entitled, Where Are We Now? The motivation for this article was that more than five years after passage of the 21st Century Cures Act, mentioned earlier, the terms "real-world data" and "real-world evidence" were being used inconsistently and interchangeably. The content of the article, as you see: address two common misconceptions and provided conceptual overview. Then the last 3 of 5 items are grayed out because I've already discussed FDA

demonstration projects and guidance, et cetera. So, next slide, please.

I just want to offer two misconceptions and hope that this discussion helps to clarify them. First is that real-world data and real-world evidence are new concepts. As my historical context showed, in reality, sources of data and types of study design haven't fundamentally changed. What has changed is access to more detailed clinical data is evolving and the data are becoming more relevant and reliable as the community works on improving the quality.

The second misconception is that there's a simple dichotomy of randomized trials versus observational studies. In reality, trials are defined by assignment of treatment, but single-arm trials face challenges similar to the challenges of observational studies in determining whether differences in clinical outcomes represent actual treatment effects when randomization isn't involved.

The next slide follows from that second

misconception. I won't spend too much time on this, but I'll go from top to bottom: randomized interventional, non-randomized interventional, and non-randomized, non-interventional studies is a little bit of jargon, but it does divide the landscape into three general categories. The next row down, we see traditional randomized trials, trials in practice settings, externally-controlled trials, and observational studies.

The main take-home message comes from the bottom of that central figure, where there's a bracket saying, "generation of real-world evidence," but it's fine if we use real-world data to plan a clinical trial, but that doesn't give us any real-world data in terms of the drug outcome association that finds patients or it identifies sites. So just in terms of what Congress mandated us to do and what we're obligated to report, it's really the 3 of 4 columns to the right where real-world evidence is generated, and that involves an increasing reliance on real-world data. Next slide, please.

When we do get real-world evidence, what does FDA do? This is a very high-level overview of our approach. We ask questions related to these three domains: first, whether the real-world data are fit for use, and that is reliable and relevant; second, whether the study design can provide adequate scientific evidence to answer the question; and third, whether the study conduct meets FDA regulatory requirements. These questions actually could apply to clinical trials, but in a different way, so we often don't need to approach it quite the same way, but for real-world evidence studies, it's a different matter. Next slide, please.

Here's an example of how we applied our approach in terms of a new indication for Prograf, tacrolimus, based on real-world evidence. The drug had been approved for the prophylaxis of organ rejection in patients receiving liver and, later, kidney and heart transplants, based on traditional randomized trial evidence, and the drug was used widely in clinical care.

RCTs were not done, at least not for FDA purposes for lung transplant for various reasons, but the sponsor submitted a supplemental new drug application to FDA with a non-interventional, so-called RWE study. The data and design were evaluated according to the standards I mentioned, and here's, long story short, the approval for this drug in preventing rejection or death for lung transplant in July of 2021. Next slide.

The reason why this worked was that the U.S. Scientific Registry of Transplant Recipients data had information on all lung transplants in the U.S. during that indicated time period. Not only was it generalizable, but the data were the same quality that we would have expected from a clinical trial arm. The non-interventional observational treatment arm was compared to historical controls, and the analysis plan and the patient level data were provided to FDA.

FDA determined that this non-interventional study was adequate and well controlled, our highest evidence bar, and I should note, however, that the

outcomes of organ rejection and death are virtually certain to occur without therapy, so the dramatic effect of treatment helps to preclude bias as an explanation of results; another way to say this is not that this was easy, but this should not be viewed as an easy way to get a drug approval. Next slide, please.

On the flip side, that was a success story. This slide is a compilation of what has gone wrong across a multitude of submissions in the three categories of data design and conduct: issues related to reliability and relevance; the need for linkage that might not exist; missing or mistimed data, mistimed being if you're not in a trial, you might not get data at the intervals that a study is hoping for; and then sometimes endpoints are the problem.

We don't have time, and this is getting technical, but threat of residual confounding; problems with the index or zero time; or the use of an inappropriate comparator in that second category. And then in terms of the conduct, we

need to be sure that the protocol was prespecified, and we also have issues related to FDA inspection that time doesn't allow discussion of. Ok, next.

As I wrap up, in summary, big data contributed to changes in how evidence generation is approached and described, and research methods are indeed also evolving. I hope I've been able to show that FDA guidance and related efforts, along with the important efforts of other stakeholders, are addressing current challenges in using real-world data and evidence so that we can improve our ability to promote the public health with drug development. In this process, we will maintain evidentiary standards while considering real-world data and real-world evidence for regulatory decision making. Next.

There are too many people to thank, but this slide is a partial list, and the last slide is an email address if we don't have time for everyone's questions to be answered; or going forward, if questions about real-world data or real-world evidence come to mind, please don't hesitate to use

this general mailbox. Thank you very much.

DR. WINIECKI: Thank you so much, DR. Concato.

I want to keep us rolling along because we have a jam-packed agenda today, and I want to make sure that we have time for the Q&A at the panel session at the end.

Our next speaker is Dr. Ramona Walls. She is the Executive Director of Data Science at the Critical Path Institute, and she has published over 50 peer-reviewed papers in incredibly diverse fields: rare diseases; environmental health; evolution; biodiversity; sustainability; and space situational awareness.

In her current role, she oversees multiple efforts, including the development of C-Path's Data and Analytics Platform; expansion and modernization of C-Path's data integration pipeline, which encompasses new data types; and the development of a rare disease knowledge graph. She's going to highlight today some challenges related to siloed and non-standard data and how to organize data to

increase its utility.

Dr. Walls?

Presentation - Ramona Walls

DR. WALLS: Thank you so much, Dr. Winiecki.

Yes, as mentioned, I'm going to highlight some of the recent developments in data science and data management taking place at the Critical Path Institute, but I'll also give you a little introduction to C-Path for those of you that might not be familiar with it. Next slide, please.

I don't think I need to tell anyone on this presentation that rare disease data are rare. We know that because the patients are rare, and as a result, progress towards therapy for rare disease patients is hampered because we don't really understand what rare diseases are, what their natural history are, and what might work as treatments.

Nonetheless, there is potentially a lot of useful data out there, particularly around real-world data. As we just heard, there are electronic health records, patient-reported registries, but

there are also more traditional data sources like clinical natural history studies, and of course data from past clinical trials, and those really high-quality data sources like clinical trials are important for helping us to understand the potentially messier, less-controlled data from real-world data.

So that's a lot of what we focus on at C-Path, is integrating those different data types and making them more useful. Unfortunately, for us, and for the patients, many of those data sources that we do have access to are siloed. They're non-standardized and sometimes they're not usable due to data quality issues, which is a real waste when you get data, and you someone's worked so hard to collect it, and you really want to make use of it. Next slide, please.

Let me first highlight some of the challenges that we see [inaudible - audio gap] not being able to understand necessarily what the different variables in a data source are because they've not been standardized, or mapped to a

standard vocabulary, or there are no dictionaries. Often, even with the best intentions of the data collectors, standards may not cover all of the variables or the different pieces of data described in data sets, for rare data particularly.

Secondly, the data sources are often siloed in that they may not be accessible. They come in different formats. They use different standards that make them challenging to integrate them. And finally, because there are such small patient populations in rare diseases, those patient populations are often distributed among multiple data sources. So it might be that there are several groups collecting data or they might visit multiple medical centers, and if their data are distributed among those different sources without a reliable method for uniquely identifying the patients, it makes it very difficult to gather longitudinal data on patients, which is extremely valuable.

So how do we start to untangle this giant ball of string, which is patient data, and real-

world data, and clinical data, and put it all together into something useful? Next slide, please. That's really the focus of what we do at the Critical Path Institute, or known as C-Path. What is C-Path and what do we do? Next slide, please.

Our mission at C-Path is to act as a catalyst for innovation that accelerates the path to a healthier world, and our vision is to be an indispensable partner of excellence in medical product development worldwide, shaping innovative, scientific, and regulatory pathways to accelerate the delivery of therapies for patients in need. Next slide, please.

We do this through a number of different methods and using a number of core competencies. The first step at C-Path is to identify and unmet medical need. That might come internally. That might come to us through a community group. That might come to us from information from a regulatory agency, but once we've identified an unmet need in medical product development, we do start to then apply our core competencies. Those include data

management and standards; the development of biomarkers; predictive modeling and analytics; clinical outcomes assessments; and regulatory and development science. Through those, we combine all of those competencies. We work as a team. We have multiple teams that we all work together to develop drug development tools and other solutions. Next slide, please.

More specifically how do we do that? The key is that we want to act as a trusted neutral third party. We are a non-profit organization. We have a lot of regulatory experience, a lot of data science experience, and a lot of modeling experience, but we do it as a neutral third party that is open to anyone who needs to use our tools.

We develop public-private partnerships. We are funded in large part through the U.S. FDA, but we also have these public-private partnerships with industry, where we convene scientific consortia with our partnerships among industry, academic, and government agencies that share data and expertise to help us basically do the best science, gain the

broadest experience, build an active consensus, and share the risks and the costs for developing tools that might not be feasible to do for any one sponsor. Through our neutral convener status, we are able to enable iterative development with regulatory agencies like the FDA, EMA, and PMDA to participate in new methods and assess the safety and efficacy of different medical products. Next slide, please.

A little bit more specific workflow of how we do that with the overall workflow within C-Path, so why do we do it? First, we know that not every drug works for every patient, so you need to target the right patients, and that's really about data. We look at the patients and try to understand their population.

Who is doing this? This is a combination of researchers both inside and outside of C-Path, working with regulators, working with groups, be they academic or industry, that are conducting clinical trials, and working very closely with advocacy groups to understand the patient voice in

the process as well. We gather data from past clinical trials. Tradition, we've relied on data from past clinical trials, but more and more we're also including real-world data.

We spend a lot of time standardizing and integrating data to different models. Those include CDISC standards like SDTM, OMOP, or Observational Medical Outcomes Partnership, using ontologies. Once those data are standardized and integrated, we're able to put them into informative models. That's where our quantitative medicine comes in to start to work with our different consortia to develop tools.

What do those models do? They can identify biomarkers. They can be used for clinical trial enrichment, developing disease progression models, and again, we work to get those models and tools validated and approved, or endorsed, by regulatory agencies so that people that want to use them know that they're trustworthy. We hope that those result in the right target, the right drug, at the right time, and for the right patient. That's

really our end goal. Next slide, please.

As I mentioned, we've got this whole workflow that includes a lot of different efforts along the pipeline, from data sciences, data management, through quantitative medicine, and through the activities of our different consortia and partnerships, and through our regulatory science team. In this presentation, I'm going to focus on the data science piece of that. That's the first piece that is the bedrock of it, that gets the data and puts it together into a useful format.

You'll hear later from one of our consortium directors, Sorin Fedeles, about some of the work that one of our consortia is doing. But let me focus here, again, on what are we doing in data science, and how we're trying to advance the field of data science, particularly for medical product development. Next slide, please.

Within C-Path, one of the key departments, the department of which I am an executive director, is the Data Collaboration Center, or DCC, and the

DCC's mission is to enable multiple organizations to work together in a neutral setting and share data to maximize its value for medical product development and regulatory decision making. But we do that first through the creation and administration of data storage and collaboration platforms and through the planning and execution of multi-source data standardization and aggregation methods. We like to maximize the fairness of data by developing and integrating standards and semantic models; developing tools for consumption of sharing of data; performing data transformations that increase data accessibilities; and by performing analyses that transform data into information.

We are not the data science team that's turning data into models, but we're basically turning data into information that's useful for models and for all of the other tools. It's really important to us that we use robust repeatable processes to ensure data integrity, security, and protect patient privacy.

Within the DCC, there are four core teams, the Data Management team, who does all of the hands-on work of data acquisition, curation, and integration; the Data Science and Ontologies team that's responsible for semantic data modeling, metadata annotation, analytics tools and statistical modeling; our Data Platform team, which is really the sort of physical, or I guess more virtual, infrastructure, designing, and developing, and testing our different platforms and products and supporting Cloud infrastructure and data security; and of course the very important Operations team that keeps us all running and functional. Next slide, please.

So I threw this word in the last slide about maximizing the fairness of data, and I realized I need to explain what that means because there may be people on this who have not heard the term "fair data principles" yet. FAIR stands for findable, accessible, interoperable, and reusable. If you're on this call, that means you probably care about data, therefore I think that you should know about

the FAIR data principles. If you haven't seen the paper yet, there's a link here. It's a short paper in Nature from 2016 by Wilkinson, et al. that highlights what the FAIR principles are and how you can achieve them.

One of the key aspects of FAIR data principles is that they're really applying to both human and machine-driven processes. Humans have an innate understanding of what data mean, of the semantics of data as it were, but humans can't operate at scale, and they make mistakes. There are errors with machines, but largely machines are able to operate at scale with much less error, and particularly in this age of big data, we need solutions that scale.

So the FAIR principles describe how you can collect, manage, and share your data in a way that is scalable, repeatable, and reducible to make your data findable, accessible, interoperable, and reusable. They really come down to principles around meta-data, metadata, metadata, identifiers, and sharing standardized protocols and best

practices around sharing and storing data. So if you haven't seen them yet, please go out and read the paper on the FAIR data principles, and embrace them, and make them part of your everyday practice. Next slide, please.

How are we doing that within C-Path? We have an approach to data management that's a multi-step process. It begins with a data contribution agreement, so we want to be very clear that we are not the owners of the data; we're merely custodians of the data. It is the organizations that are contributing the data to us that maintain ownership, and they in turn are behaving as custodians for the patients and individual people about whom the data is.

Once the data contribution agreement is signed, the data are transferred to us through a secure link. We generally only accept anonymized data. We are not storing PHI, personal health information, within C-Path; however, with the growth of electronic health records and other real-world data, we have started to make occasional

exceptions where we can work with PHI, but we're generally using anonymized data. So we can also work with our data contributors to help them understand what they need to do to anonymize their data.

Once we get the data, we curate it, we standardize it, and we annotate it with terminologies and with links to other data. This blue arrow here shows an important step, that we provide feedback to the contributors. When we find problems with the data, we report those to the contributors. Now, if it's a past clinical trial, there's not really much that can change about it, but if we're working, for example, with a registry, we want to work with them and give them feedback on how they can improve their data collection processes going forward.

Once we've got the data in-house, and we've standardized it and curated it, we integrate it into different databases as part of our data-sharing platform, where it's available to approved researchers -- those may be internal or

external -- to extract data, and analyze the data, and combine it potentially with their own data for additional analyses.

Over the past few years, we've been making a lot of advances and innovations at each of these departments, so I'm going to just step through each of these steps. I'm going to walk through them and talk about some of the innovations that we've been applying at each step. Next slide, please.

When it comes to data contribution agreements, or DCAs, we've been working on standardizing those rather than having an individual data contribution agreement for each data source. We've been trying to have a small subset of them for different uses. That makes it much easier for us to manage the data and for us to explain to potential re-users of data what those conditions are on the data. We're also moving towards machine-readable data contribution agreements, which, again, make it easier for us to manage the data and ensure that we're being compliant with the terms of the DCA when we do

share it. Next slide, please.

For transferring, we've moved largely to a Cloud-based system for all of our data, so we use common Cloud platforms, your AWS -- we're not using Google Cloud -- and no endorsement of any of these systems is implied here; we just use different ones. But why is this important? One is for security reasons. We now have a secure method so that contributors can upload their data directly to the Cloud for us, so it never has to be on anybody's personal computer.

As I mentioned, because of the growth of real-world data, we're starting to offer some anonymization services through the Cloud, and we've been really focusing, as much as the world has, on federated access and federated analyses of data. There are a lot of challenges around that, which aren't really the topic of this presentation but a recognition that sometimes data need to stay where they are. It doesn't make sense to move really large data sets around, so we need to go out and move our analyses to the data, and we've been

working on methods for that within C-Path. Next slide, please.

In curation, standardization, and annotation, we've seen a lot of changes within C-Path over the past few years. We've developed a process that we call responsive curation, and that has to do with, really, rather than a slow process where all the data come in, it sits on our data store. Our data managers take it and spend six months to a year curating and getting everything beautiful before we can do analysis on it. We do the curation more in a step-wise process, so groups will come to us and say these are the variables that are most important or these are the data sets, and we focus on curating pieces of the data set at a time as is required, so we can prioritize curation to the data sets that are the most valuable and the most in demand.

We've also moved away from simply using the CDISC standards. We continue to use those, though; they're very important. But with the advent of real-world data, we've also adopted the OMOP

standards, the OMOP Common Data Model, which is the Observational Medical Outcomes Partnership. We're also starting to use ontologies such as OBO Foundry ontologies like the human phenotype ontology, which are also being incorporated within the OMOP Common Data Model vocabularies.

We started using scriptings and automations to try to speed up the curation process as much as possible, and we're developing an ontology and a knowledge graph that allow us to really integrate data and make additional inferences from data in a much more robust fashion. Next slide, please.

Within the integration and data-sharing platform, we do have a new platform specifically for rare diseases called the RDCA-DAP or Rare Disease Cures Accelerator-Data and Analytics Platform. That platform has advanced search discovery, and visualization, and subsetting tooling available, where once you've requested access, you can go in and preview what data are available, do queries on it to see how many missing subjects are there for different variables; that sort of piece, to find

out if the data are valuable before you go through the request process.

Once you have requested access to the data, you can move it into a platform where there. Sorry, I'm getting ahead to the next one. Let me talk about this one about data sharing. We have access in terms of sharing. Rather than having to share an entire data set, an aggregated data set, we can share different pieces, so we have these fine-grained controls within there. Again, similar to the data contribution agreements, we are trying to standardize our data use agreements to make it much clearer and easier for users to understand what their obligations are when they are requesting access to this data, and what they have to report, and how to use it appropriately, while protecting patient privacy and intellectual property as well. Next slide, please.

As part of the platform, we also have a workspace. There are places where you can come and do the work once you've requested access to it. You can move the data into a workspace that has

built-in tooling for analysts like data previewing using R, SQL, and virtual machines for doing customized analysis. There is a lot of enhanced security on our platform that includes logging of all activities; TFAs, two-factor authentication; and restriction of downloads. You need to request permission to download data, and that will, again, reflect what was signed in the data use and data contribution agreements.

You can also share. It's a collaborative platform, so you can share your analyses with other collaborators and with regulators. If you've done your work in the platform, if you've developed a tool and you want to share it with the FDA, you can invite them there to come directly to the platform and do the review of your tool and the data right there, and you can also bring your own data. If you have private data that you want to add to public data sources, that's possible. Next slide, please.

Here's just a screenshot preview of the data discovery part of our platform, of RDCA-DAP, what's

called FAIR Data Services, and there's the use of the term "FAIR" again because it is trying to make data fair. Through the FAIR data services platform, you can come in. You can do a search. You can browse the different data sets. You can request access to them. You can view the data dictionaries to see what data are there, et cetera. Once you've requested access -- next slide, please -- you can move the data into a workspace, and workspaces are where you can do the actual analyses. You can do previews. You can share all of the different features that I mentioned in the last slide, so these are the tools that are available.

Now, this is right now called the RDCA-DAP, the Rare Disease Cures Accelerator-Data and Analytics Platform, so it's appropriate for rare diseases. But I'll mention that we are moving this to become the C-Path DAP, the C-Path Data and Analytics Platform. So it will not only house our rare disease data; it will ultimately house most, if not all, C-Path data within this platform, and

we think the security and functional advantages of this platform are so great, that it's worth moving that into making this our main platform. Next slide, please.

Just to wrap up this section on innovations with a little piece about what we're doing in terms of data standardization, as I mentioned, we're now using both the OMOP Common Data Model, as well as the CDISC SDTM data model. They both have their advantages and disadvantages for different situations, so we are continuing to use both of them.

SDTM is really crucial. If we're only integrating clinical trial data that's already in that model, it's really the best choice. On the other hand, if we're using real-world data and we need to use a long-tail registry data or very large EHR data, then OMOP tends to work better. OMOP conveniently uses standardized vocabularies from the Unified Medical Language System, UMLS, like SNOMED, LOINC, RXNORM, and CDISC on the other hand is already linked to NCIT, the National Cancer

Institute Thesaurus, so there are big differences in their vocabularies. And again, there's no perfect biomedical vocabulary out there yet. We do a lot of work to map across all of these different standards and vocabularies, and that's where ontologies come in.

We are using OBO ontologies. OBO stands for the Open Biological and Biomedical Ontologies Foundry, or OBO Foundry, which are a set of very semantically enriched ontologies. Unlike ontologies, say, in SNOMED, which has hierarchical structure and some relationships across different pieces of the ontology, the OBO Foundry tends to be more robust in explaining exactly what a term is and how they're defined. That allows us to encode additional information within those levels and do a deeper level of integration than might be possible using simply the OMOP standard vocabularies or NCIT.

What we're doing with those within rare diseases is building a knowledge graph, and that knowledge graph is quite different from others.

There are a number of knowledge graphs out there, and some really good ones, but what we're doing is integrating many of those existing knowledge sources with patient-level data because we have access to individual patient data in C-Path, and we're making sure that we're interoperable with those other sources like Orphanet or the Monarch knowledge graph, and the European Joint Programme on Rare Diseases.

Again, the main focus of this talk is not knowledge graphs, but since it might be a new topic to many of you, let's go to the next slide, and I'll just give you a quick preview of what a knowledge graph is. A knowledge graph is essentially combining the data plus the ontology. So the ontologies provide a model of experts understanding of what things mean in the real world, and the data are actual instances of patients who have these diseases.

In this particular case, if you look on the bottom right with all the blues, there's the clinical data condition occurrences. That tan dot

in the middle is the class for, in this case, Friedreich's ataxia, and then we've got all of the different individual observations of patients with Friedreich's ataxia in blue around that. But because that Friedreich's ataxia disease is linked through the ontologies to all this other knowledge, it connects up to cross species knowledge about gene expression that might control ataxia's morphological information about how body functions and parts relate to one another, and other phenotypes that are specific to that disease and might relate to other diseases.

So basically, the knowledge graph allows us to connect patients to the larger world of biomedical knowledge that's out there, and make some inferences about what patients might be similar based on their phenotypes or their genotypes. How might the phenotypes of one disease relate to another disease? How might we understand some of the preclinical work that's done in model organisms? How could that inform development of drugs or clinical trials within humans, for

example? That's just the highlights of some of the work that we've been doing within the Data Collaboration Center at C-Path. Next slide, please. This is what we do. We take this data and we try to make it useful as possible. What can you as data contributors, people who are collecting data and working with patients, do to help make this whole landscape better and more effective? Whether you're a small or a large generator of data, this can apply to you, hopefully. Next slide, please.

First is sharing data in an appropriate way. I'm just going to highlight a couple of slides here from a webinar that we gave last week through the clinical research data-sharing lines, and it's based on a paper that recently came out in applied clinical trials. In this webinar, we discussed -- it was the results of a survey. I won't go into all the details of the survey, again, because you can read the paper.

Basically, it's clear that some documents need to be shared that are more important than others. I'm sorry this is a bit small, but

basically we have the ADaM Data Set, the SDTM, the Data Dictionary, the Digital Specifications, and the Study Protocol. Over 80 percent of patients said that those were important, and all of these supplemental documents, for all of them, over 90 percent of patients said that they were mandatory, or important, or at least useful. People who are reusing the data need the supplementary documents, so if you're going to share your data, please be sure to share the information that allows others to understand what your data mean.

A particularly important piece is the Variable-Level Transformation Report. When data are anonymized and shared, transformations happen. If others don't know how you transform your data, it's very difficult for them to then go in and reuse it. Next slide.

But ironically, even though we know those documents are important - and- this is only for companies; this is not registries or academic institutions. Companies are not necessarily

sharing that important information.

Tier 1, or large companies, which was over, I think, 25,000 employees, are consistently sharing the required document, probably because they have the resources and larger data-sharing teams to do that, but as you move into smaller companies -- and we're pretty sure that we know from experience, this is also true for academic institutions,-- those documents are not being shared. So there's a real mismatch here between what's required of people who are using the data and what companies are willing to share. So in other words, we suspect a lot of people are just checking off the box saying, "Yes, I shared my data," but they haven't really done the due diligence to share everything that's necessary to make that data useful. So what should you do? Next slide, please. I'll wrap up with this. Follow FAIR data principles, know what they are, and try to follow them. Make sure that you ensure proper anonymization and include your anonymization report when you share your data. Where possible, use

standardized terminology and data models. OMOP and SDTM are two good ones, but they're not the only ones.

Use standardized vocabularies like the UMLS, use the NCIT, the common data elements from NIH. Use ontologies like the Human Phenotype Ontology to describe phenotypes. Phenotype is a very broad term here. That includes everything from hair color, to organ function, to clinical outcome assessments of patient performance.

Following consistent data protection practices from year to year, I know that's not always possible for smaller groups because you collect data for a year, and then you learn what's more important the next year, and you improve it, and then you learn more, and then you improve it. But because longitudinal data are so important, the more that you can aim for backwards compatibility, at least with your data, the more valuable your data will be.

Especially share your dictionaries, share your protocols, share the other supplemental

documents, and work with those who are going to reuse your data to make them understandable. Realize that as a data sharer, you are probably also a data re-user. Most people that share data also reuse data, so be a good player, be a productive part of the ecosystem, and make sure that you're not just checking the box when you share your data, but you're contributing data that's actually valuable and doing the most service to your patients about whom that data are collected.

With that, I believe that's my last slide. Next slide, please. Thank you very much, and I'll pass it off to the next speaker.

DR. WINIECKI: Thank you, Dr. Walls.

Now we're going to move to our third talk. Our speaker is Vanessa Vogel-Farley. She is the Senior Director of Research and Data Analytics at Global Genes and the principal investigator for the RARE-X Data Collection Platform. She possesses 20 years of experience in data collection methods, as well as expertise in non-profit and research

operations, patient advocacy and support, and non-profit management. Her talk today covers expansive topics, from privacy and data governance, to how to organize and share data for the maximum benefit of all shareholders.

Vanessa?

Presentation - Vanessa Vogel-Farley

MS. VOGEL-FARLEY: Thank you so much for having me. My name is Vanessa Vogel-Farley, and I serve as the senior director of Research and Data Analytics for Global Genes and their RARE-X program. To change the world for rare disease patients globally, we must think differently. One of those ways is by increasing the speed and productivity of innovation in rare diseases by increasing the collection and access of structured and standardized patient data, which is what I'll be talking about today. I actually want to rename my talk, basically, to Make Ramona's Job Easier. That's what I should rename it. Next slide, please.

The speed and productivity of innovation in

a rare disease is often limited by cost and lack of access to standardized, structured, and available patient data, which you've heard from the two previous speakers; or data exists in silos and is unavailable for open research; or data is not structured and standardized in a format that's useful to research or patient communities; or data just doesn't exist yet since many patient communities are too young or don't have the resources to connect data for research towards treatment development. These are the areas that hold promise of unlocking data in various ways. Next slide, please.

Patient organizations in the rare space often start from the ground up, forming registries for their communities to gather the much needed data that we've been hearing about, but how do we go from registries to real-world data and show what patient-powered registries can really enable, all the way to supporting regulatory requirements? Next slide, please.

We're living in a world where patients and

patient advocates have more opportunities than ever for helping to overcome some of the data collection challenges that drive biopharma, where patient groups are partnering effectively with biopharma, governmental regulators, and goal networks, and they're becoming investors in that space, and some are even becoming biotech entrepreneurs. I'm going to focus today on the patients as research and development partners and drivers and how can patients and advocacy groups support the collection of patient-reported outcome data in a way that can actually be valued and used. Next slide, please.

The process of data collection and research in clinical trials starts with the process of consent -- so I'm really going to start from the bottom -- and ensuring data is accessible as possible with the goal that accessibility extends post the initial intended purpose needed to decrease the time to new treatments for rare diseases. Next slide, please.

Consents and protocols should include language and supported patient-focused data

governance and standardization language for broad data usage. What this means is those who are collecting data in this space, while you might not have started in this manner -- meaning consents or governance protocols might not allow for data sharing in a more robust way, -- the time is now. There's no time like the present to review and evaluate your existing consents and protocols to create enabling data-sharing language and to add data management procedures and recommendations with inclusion and usage of this data collected post the original intention of the data. There are also opportunities to create more robust data on ecosystems around rare disease communities by enabling this. Next slide, please.

So how we do this at RARE-X is we actually go beyond the single-informed consent for data sharing. This is an example of how collecting data use preferences in a direct efficient manner so that it can be used in a machine-readable manner, sort of like make Ramona's job a little bit easier. So we're asking the patients themselves where they

want the data to be shared from the point of inclusion and any data collection efforts that we're doing. Next slide, please.

To leverage data use ontology, I want to talk a little bit about this as well. Ontology is general ways of labeling data, a variable or something that's coming into your system, that creates a meta-data or a meta-item. How we use those for data sharing, we educate the patients in a two-prong approach when it comes to empowering patients to share their data. In our case, there's a presentation of the data use options, which are the ontologies. You use our GA4GH data-sharing preferences that are shown as part of the consent process that's direct to the patient and what we call the Data Sharing Preference Survey, where there's a separation of the represented data-use ontologies to enable the patient to review those independent of the rest of the study consent.

So it's outside of what this study talks about, it's outside of the data you're collecting, and it's really just saying, okay, we have this

data and we are consenting to have that data collected, but now, where do you want your data to be shared after the intended use? And they're able to review those potential data-sharing options multiple times and update those outside of the consent document itself. So over time when they're participating in longitudinal data studies, they can update them, depending upon the data sharing opportunities out there. Next slide, please.

We use data-use ontologies, which is a structured vocabulary of standard, human, and machine-readable use terms that have been adapted in a patient-friendly manner. I know this is really small, but what we did is we went into GA4GH data-sharing ontologies and made it more patient friendly, the way that we describe the types of data collection and data usage that are out there, and made sure that they could understand it in a very patient-friendly way, and also made it more specific to patient data. There's a lot of things in GA4GH that's from genomics data and large data usages, so we really made the ones that were more

specific to the patients available to them. Next slide, please.

So outside of consent, what are the next steps for using standards at the time of data collection? Basically, it's how we make efforts like C-Path's efforts more robust and easier. Next slide, please.

But when it comes to data collection models in the rare space, since there are more than 10,000-plus rare diseases, we need to take into account the splitting and lumping that are needed to address as many patients as possible.

For example, we know that in the rare space there are N of 1's. There are individuals or the undiagnosed population where they're still on their diagnostic journey, or we have patient communities that vary from a couple patients all over the world to really large patient communities that are in the rare space. Then we have the disease consortia, where they're based upon body system or symptoms that bring together several disease communities around one symptom, and usually towards better drug

treatments or drug interventions because they can address that symptom rather than necessarily the disease as a whole.

There are data collection challenges with each one of these, but starting with the data collection model based on standards, we have the ability to ensure that any data collected is able to be used in a data ecosystem, similar to what Ramona was talking about, more quickly than those that are not. Next slide.

To meet as many stakeholder needs as possible, the standards and guidance that are consulted by RARE-X are the ones that you've heard about, the alphabet soup, and I know that somebody in the chat actually asked for a definition of a lot of the alphabet soup that we've been talking about. CDISC, Human Phenotype Ontology; the NIH Metathesaurus; the Common Data Elements Repository; PhenX; LOINC; SNOMED; Orphanet, ICD codes are all part of those, but also guidances that are put forward by regulatory bodies like FDA, which was presented earlier. Those links are in the chat,

and they will also be in our slides, so make sure to look at those, and NCATS guidances; the scientific community; industry partners; and in our case in the rare space, guidance from patients, too.

Data collection in this space, when you're looking at small n's or you're looking at communities that are really spaced out, guidance from patients is really needed to make sure that your data collection is able to be robust and maintained over time, especially when it's based on standards. Next slide, please.

The application of these data standards and data models to provide infrastructure to support comprehensive data for analysis, we need to gather precise data, map it to the ontologies, and layer it with other data sources, and share it, really, to make sure that that's data getting out there. Starting with a general core in RARE-X is an example of how we collect standardized data and how we create our data models. We start with a general core, where it's a head-to-toe survey, where every

patient that comes in gets it, and lets us know what part of the data model they're going to plug into, what's being affected in the disorder that really means something to them, and what they want to give more data about.

Enabling disease core by domains, where these are HPO mapped domain-specific data, and layering them on supplemental disease data that can be detailed or more specific to that disease, and then integrating other data sources like EMR and EHR, which were talked about, and some clinical reports, and maybe some custom curation forms around genetics or labs, or those sorts of things, while always allowing the flexibility for exploratory data collection; since in the rare space, we need to acknowledge that there are areas with standards that just don't exist yet, and we really need to make sure that we're addressing those in capturing data around those in these patient communities, as well as making sure that as we're capturing that data in these more structured ways, that we can move towards making new standards

that meet the rare disease needs. Next slide, please.

Just as a little bit of definition of our general core, a general core for us is a data element that can be consistently collected across all disease communities in all studies or therapeutic area. A disease core element is a data element specific to a therapeutic area or specific disease constellation of central modalities, like you're looking at a therapeutic area of epilepsy, but lots of diseases have epilepsy, so that's a disease core where it's one of the most prevalent symptoms in that space, so that's one of our questionnaires around that.

Then there's supplemental or custom surveys, where our data element is commonly collected in clinical research studies, but whose relevance depends upon the study design and the type of research steps involved. This is kind of getting back to the real-world evidence and real-world data applicability, and these can be developed on a case-by-case basis, based on standards and

ontologies towards robust implementation in that larger data ecosystem. Next slide, please.

One effect of data models used in this manner is the investigation of disease overlaps, and symptoms and disease biology is unlocked. Here's an example of our three semi-different disorders with similar mechanisms of being an ion channel disorder are able to be compared with their similarities and their differences. These sorts of analyses can bring a core of targets that have never been identified before in drug development.

In the rare space, this is so important because when it comes down to it, yes, we are rare, but there are so many things that we do overlap in terms of symptomatology and also targets when it comes to drugs. So why not actually lump when we can and split when we need to when it comes to these sorts of things? And when you're basing your data collection on standards and you're basing your data collection on really robust governance, this enables that really, really well. Next slide, please.

To gather data to facilitate each of these data elements that we've been talking about, we need to do that in a domain-based standardization module with machine-readable ontologies where we can move it through a system, like we've been talking about with C-Path. Here's a quick sample of the domains we collect currently on the RARE-X platform, as well as some of our domain expansion prioritizations -- next slide, please -- like how to prioritize, especially when you're going into this space where you're saying I'm a patient community leader or I'm a researcher entering into some of the rare disease spaces. How do you prioritize what you're going to collect and how do you structure your data model?

Well, you turn to the experts, and that includes patients. In order to prioritize any data collection effort that we do for research-grade and comparable data, we establish multidisciplinary expert working groups for each of the domains. Some of them might overlap and some of them might not, but as you can see here, they represent

pharma, they represent the patient groups, they represent clinicians, and they represent academics, to make sure that we're bringing forward the right symptom domains, and landscaping what's out there and what's going on in the space right now, rather than relying on studies that have been done decades and decades ago.

Then categorizing those patient-reported outcome measures or those clinical outcome measures that really need to be brought forward for these community groups, and then deeply review and discuss those measures to narrow them down. What's too long for these patients to sit down and do it at one time? How do we kind of layer those aspects where this is a good layer that we can jump off and branch to get more data in more standardized areas? Then confirm the final measures to the level of data collection being focused on, depending upon what the domain is.

Then we go through all the paperwork of licensing and technical implementation, which I'm sort of glossing over, but that ends up being a

really, really big deal when you're coming to the space of standardized data collection. When you're using license-validated measures or using a survey that might be based on ontologies, the technical implementation of licensing is really, really important in that space. Next slide.

One of the questions answered and posed was how to best use the data from natural history studies for rare diseases? Up until now, the domains that I've been talking about are mostly patient-reported outcomes, that we bring the data collection to the patients, because at the end of the day, we know that rare disease doesn't have any borders. It doesn't have any SES regulations, and it really affects everybody. So when you're coming into the space, how do I make sure I get data collection direct to the patients where they are?

In the space of the natural history studies, in the past, you have to bring the patient to the data collection. One of the ways that we're approaching natural history studies in more of an agnostic way and gaining some traction are more

basket-style natural history studies. We hear about basket-style clinical trials, but what about basket-style natural history studies across rare diseases? Many clinical and research programs launched for multiple rare disorders are similar in phenotype, and due to the increased demand, how do we help clinicians and researchers collect the data and point of care in natural history study data?

We're in pilot phase with a clinic that has a neurogenetics focus, where clinical outcomes assessments are most applicable to the patients that have been decided and are collected as part of clinical care, where they include clinician-reported scales, clinical observation assessments, patient-reported scales, as well as the platform that's available to them via RARE-X.

The data model that was created was done based upon a working group really similar to the one that I've just described, but really bringing it down to what can you get done when they're being seen in clinic, what really makes sense when you're looking at the holistic patient, and what makes

sense that we can collect over time; so really making sure that we're addressing what is being able to be collected when a patient's being seen there, and then also additional data sources like EHR to decrease the duplicative entry of data so you're not answering a question twice, so clinicians aren't entering something in the EHR as well as in the research record, and making sure that we're bringing together those data sources on the background and leveraging the technology that exists to do that these days. Next slide, please.

I talked a little bit about validated instruments. In rare research, validated instruments sometimes become a little bit of a sticky subject. Validated instruments are also known as questionnaires, patient-reported outcomes, and clinically-reported outcomes that have been studied extensively, using specific scientific criteria and statistical methods that give us confidence that they're reliable and valid in the population used to validate the instruments.

For an example, an instrument validated to

help people with cancer may not be applicable to caregivers of rare epilepsy, just as a really random example. But there's also FDA definitions of all of these things, so when you talk about validated instruments, they're really important because we know they're valid and we know they're statistically reliable for data analysis, but -- next slide, please -- there's a catch-22 when it comes to validated instruments in rare disease. We need them for regulatory purposes -- we know this -- but they often force us to use proxy-reported outcomes when it's coming into the rare space, when the patients themselves are not able to answer for their own feelings and those sorts of things, and it results in data that may not represent what the patient is actually experiencing. It might be representative of what the clinician is seeing, or what the caregiver is seeing, but it might not actually be what the patient is seeing.

There's a need in the rare disease space when it comes to validated instruments for the

development of new ones to address these challenges, and the acceptance and qualification of appropriate instruments [inaudible - audio gap] -- in standardized data collection -- can use a question that's not validated and still be seen as compliant or ontology compliant, and could be seen as one of the ontologies, but it could be just in general standards compliant. This is a very important for rare diseases, where validated instruments tend not to hit the mark, as I just talked about. Next slide, please.

The answer is yes, but tread carefully. As we've heard from the last two speakers, when you're doing research and entering to the space, you want to be thoughtful about how you're implementing your disease or your data collection. There are many recommendations out there that will meet the requirements, but make sure you're opening up that conversation early and often.

The FDA has fantastic contacts, that when you're entering into the space, whether you be a researcher, a clinician, a patient advocacy group,

or a biopharma who's entering into a new rare disease space, reaching out to them to say this is what we're collecting, this is what we think the purpose is, and this is how we're thinking about designing these efforts, is really good, and to engage them early and often because it's really needed in this space to make sure that the communication around your data collection efforts is clear, and what you're collecting from the patients is really worth the time and the effort, so it's fit for purpose. Next slide, please.

At this point in the story, we've got data. We've got consent to collect the data. We've collected the data in a hopefully more standardized way, where the data is able to be used past its intended point. It has the ontologies to be able to move through these different data systems. But now, how do you connect the other data sources that are existing?

You might be collecting your own data in your academic environment, or your patient advocacy group, or your biopharma. We know that other data

sources exist in all these spaces, and for rare, in order to make that data ecosystem or that data map for that patient, or that patient community, we need to be able to connect these data sources. Next slide, please.

This is just an example of the way that you can interconnect and support other data. I'll focus a lot about data generation and data governance. There's data in many communities, and it is important to make sure we're able to connect towards research questions and towards clinical trial design. This includes EMRs, historical physician notes, diagnostic testing, and journey information, as well as additional studies that our advocacy groups are supporting or researchers are supporting, and that we are partnering with biopharma on.

In the last 5 to 10 years, the speed at which Cloud computing and federation of data technologies are being brought forward is so exciting, and being able to have these data sets accessible in a federated manner, or an uploaded

manner, can really unlock the potential of all these data sets. Sometimes this means being directly connected to the data. It's uploaded, and you're actually getting it out of there on a direct patient basis. Other times, the data needs to stay deidentified in some of these areas, or actually where it was, as Ramona said earlier, but could be used as a comparison or hypothesis testing analyses, especially in a rare space when you need those comparators to be able to do effective data analysis.

Sometimes when governance inhibits data access, it may be useful just to have the previous data models to determine the efficacy of that data collection effort to potentially incorporate or improve new data collection efforts, meaning that if you've collected a natural history study and you didn't use half of the data, or used 100 percent of the data, that's an amazing model that really could be implemented in different areas, especially in this space, to create robust and standardized data collection over time. The goal and the mantra, in

general, in the rare space is meeting data where it is and leveraging technology to interconnect or federate, in whatever manner we're able to, towards no data left behind. Next slide, please.

So with this growth comes the true phase shift of how we think about data management and inverting the model of data sharing towards public good for all efforts versus commercial and closed data, those silos that we talked about earlier. The opportunity for us right now is to bring researchers to the data or data to the researchers in whatever way, shape, or form we can.

RARE-X places data in the Cloud, where the data can be computed and brought together with researchers. They can collaborate. Similar to what Ramona was saying for their Cloud-based efforts for C-Path, we do something similar to RARE-X. Researchers can store the data and access a single copy of the data, and these address the concerns of lower cost, audit controls, threat detection, with the understanding that this might not meet all stakeholder needs, but the federation

of data towards discoverability is a step in the right direction for a lot of rare diseases.

With this inverted data-sharing model, it allows data sharing in an expedited manner, as well as providing a place for researchers, clinicians, and biopharma to reposit data after clinical trials or studies are completed so that data is accessible. Many years ago, NIH mandated the data for NIH-supported studies to be a repository for future research. Can you imagine the power of data from clinical trials, both successful and unsuccessful, being shared? It would improve disease understanding and protocol design in the future, and the list goes on; but most importantly, decreasing the time to new drugs and new treatments for patients. That's really what it comes down to. Next slide, please.

Our platform, in general, enables rare disease patients to share data at scale. Researchers can then analyze the data and other federated data, using integrated tools deployed within the collaborative work spaces, as well as

making data discoverable, linkable, and accessible to other researchers, clinicians, biopharma, patients, and communities. Efforts like those of RDCA-DAP are one of those things that we connect them to. So we are really proud to have a partnership with RDCA-DAP, where the data from RARE-X is consented, and that's where the patients want their data to be shared, and it's able to be shared with RDCA-DAP and all of their efforts.

We're actually working on a really nice ontology project right now, where we're mapping our ontologies that we use at data collection to the ontologies that RDCA-DAP has historically put on data after it's been sent to them, so we're really excited about that. The barriers are lower and the time to data usage is slashed. Next slide, please.

It's important to note that the stakeholder ecosystem for rare diseases is one where patients, patient advocates, or organizations are often drivers of data collection to increase visibility and knowledge about the disorders. Without their engagement, many of these communities would be left

in the dust. However, the intricacies of data collection, purpose, and usage to meet all stakeholder needs to drive that ecosystem, where each stakeholder is able to play their role, filled by well-collected and shared data, is really what we need in this space. Next slide, please.

So I mentioned RARE-X a couple times, but this is actually what we are. We're a program of Global Genes created to accelerate rare disease research treatments and cures by removing barriers for data collection and sharing. We're a platform to collect, connect, and share data. RARE-X is not a replacement for any current research or clinician-sponsored registries, but rather a prepared collaborator and partner, ready to meet data where it is and enable access in whatever way it can compliantly be used.

RARE-X recognizes there are many different places, entry points, and challenges that any one rare disease can experience, and the approach isn't necessarily linear when it comes to approaching data collection. When establishing new efforts and

improving on existing efforts, enabling data sharing via consent and standardized models where applicable can ensure that data for rare disease patients is worth their time and effort that they give to put this data in. There's never been a better time for patients, researchers, clinicians, and biopharma to partner on data collection and sharing to kick-start what needs to happen in the future for rare diseases, and we're here to help. Next slide, please.

We can provide a platform to help collect structured patient data, including these patient-reported data elements that I just talked about, but we also want to enable open science platforms to facilitate the sharing of large high-quality data sets to accelerate therapeutic research, and a full ongoing patient engagement, program management, and service to ensure participation and success for patient advocacy groups. Next slide.

So a big thank you, and happy to answer any questions. I think we're going to move on, and

I'll turn it back over.

Q&A

DR. WINIECKI: Thank you so much.

We have run over a little bit. I want to do just a bit of a concise Q&A with our three speakers. I'm going to try to throw one question to each of them, but keep in mind that if you want to address a different question that you saw pressing in the chat, in the Q&A box, feel free to do that.

John, the one that stuck out to me that I was going to throw out to you was how to leverage real-world data in rare disease clinical trials, for example, using EMR data, disease registries, and master observational trials?

DR. CONCATO: Wow. Even if we had more time, I think that's --

DR. WINIECKI: I know it's a very broad question.

DR. CONCATO: -- a broad question.

DR. WINIECKI: Take that where you want.

DR. CONCATO: Okay. The way I would frame

an answer is if we have bookends of the spoke on one side and one size fits all on the opposite end of the spectrum, I think the key aspect is to consider where one is in that regard; how much do we know from prior experience.

I think the title of these three talks together is we're improving the field. We don't know what will be the highest return on investment, but we have to be thoughtful. So it's fundamentals of data quality, appropriate study design, and regulatory context. C-Path is doing great work. You heard from Vanessa and their particular approach. I think we're seeing -- one more phrase -- a rising tide lifts all boats. So I don't think I can answer that question except on a case-by-case basis, but that's where FDA, at some point in the process, gets involved. Thank you, Scott.

DR. WINIECKI: Sure.

Vanessa, I'm going to toss this one to you. How can advocacy groups support the collection of patient-reported outcome data in a way that will

actually be valued and used?

So I take that to mean if someone is starting a data collection effort, what are some tips you would give them so that they can get the maximum use out of that data?

MS. VOGEL-FARLEY: Sure. I actually just was speaking with a patient advocate last night that started a registry, and when we talked about it, it's not as simple as saying I'm sending out questions to families about X, Y, and Z. The way that you ask your questions, the actual intention of how you're going to use that data in terms of research and analysis, needs to be thought of beforehand.

So really, when you're thinking about that, bring forward - yes-, your community's questions are great, but then meeting with a researcher or meeting with the clinicians doing research in their space to say, now, how do I make this research-grade? How do I ask the questions in a non-leading manner? How do I make sure that they are standardized or led to ontologies that might

exist in that space, or existing common data elements or variables that might exist already?

We know that NIH has a massive amount in this space that you can actually link in to, and the same thing for HPO. So really making sure that you're bringing forward what your community wants to know, but then linking up with somebody who knows research methods in that space to make sure that you're evaluating all of those needs as well.

DR. WINIECKI: I think that's excellent.

For Dr. Walls, how do you entice sponsors to donate data, either from randomized clinical trials or real-world data to C-Path, and what are the key challenges to obtaining and getting data?

DR. WALLS: It's surprising easy to entice -- well, I shouldn't say this. Our consortium directors are probably like wringing my neck right now. But we have been very successful getting sponsors to share data because in rare diseases, the research community recognizes that no one organization has enough data to develop solutions. So if you want to understand the

natural history disease, if you want to have an effective disease progression model against which you can compare your treatment, you have to collaborate. So the only way that you're going to succeed is through sharing data.

Even in more common diseases like Alzheimer's, there are many areas where there is still no treatment, and the sponsors have been working on it for decades without coming up with a solution, and they recognize and they come to us and say "if C-Path can build this collaboration." And in some cases, we do need to protect intellectual property of the sponsors. There are cases where sponsors will say, "My data can only be shared within this consortium." The other members are the only ones that can see it.

So that's important, and that does happen, but that's becoming less and less common. The data-sharing culture in the world is growing. Sponsors are recognizing the value of data sharing not only to themselves, but to the larger community, and taking part more often.

In terms of the biggest challenges in data sharing, in contrast to what I just said, there's definitely still an education piece where we need to explain to sponsors how important it is for them to share the data and the benefits that we get from that. A lot of the challenges that I see are technical around ensuring data are properly anonymized, understanding what the data mean, how we reuse the data, and all of the pieces that Vanessa just talked about in her wonderful presentation. If we can solve all those challenges and do everything Vanessa just said -- please -- with data sharing, my job will be much, much easier, so thanks, Vanessa.

DR. WINIECKI: No, I think it's interesting. The devil is always in the details. Collecting data may not be terribly hard to do. You can just set up an Excel spreadsheet, or whatever, and start collecting data, no matter what you are talking about. But when you are talking about integrating data and organizing data and merging data, it becomes incredibly complex very quickly.

Just in a minute or so, do any of the panelists have any other thoughts or comments that they want to throw out before we take a brief break before Session 2?

(No response.)

DR. WINIECKI: Okay.

Well, in that case, I want to thank John Concato, Ramona Walls, Vanessa Vogel-Farley, and Dr. Kerry Jo Lee for contributing to this session. We'll take a brief break, and we'll be back in about five minutes for Session 2. Thank you, everyone.

(Whereupon, at 10:32 a.m., a recess was taken, and workshop resumed at 10:45 a.m.)

DR. LEE: Hello, everyone. I'd like to welcome you back to our second session for day 1. This has been a wonderful morning, and thank you all for all of your incredible engagement. We've really appreciated the questions, and tried to get through as many of them as we possibly could.

I'm just going to introduce our second session, which is going to be moderated by

Dr. Christine Nguyen. She is the Deputy Director of the Office of Rare Diseases, Pediatrics, Urologic, and Reproductive Medicine in the Center for Drug Evaluation and Research, in the Office of New Drugs at the FDA. Dr. Nguyen joined the FDA in 2005, and in her current role, she provides important leadership to scientific, clinical, regulatory, and policy considerations related to the treatment of inborn errors of metabolism, including lysosomal storage disorders, organic acid disorders, and amino acid metabolism disorders.

She has served in several leadership roles prior to her current one at the FDA, including being the former division director in what is now the Division of Urology, Obstetrics, and Gynecology within the Office of New Drugs, and we are very excited to have you here to moderate the second session.

Thank you, Dr. Nguyen. I'll turn it over to you to introduce the session and the first speaker.

Session 2 - Christine Nguyen

DR. NGUYEN: Great. Thank you so much,

Kerry Jo.

Good morning. I'm Christine Nguyen, and I'm very excited for our workshop today, and you can see all the topics that will be covered that's so applicable to what we do at FDA every day.

Our first presenter, Dr. Sorin Fedeles, is the Executive Director of the Polycystic Kidney Disease Outcomes Consortium at the Critical Path Institute, and there he oversees the strategic vision, management, and activities of collaborative research endeavors with various stakeholders. His work and leadership related to the therapeutic development for the treatment of autosomal dominant polycystic kidney disease, which is the most common genetic cause of end-stage renal disease, has spanned over his career, both at C-Path, and while also in faculty at Yale University School of Medicine, where he remains affiliated as an assistant professor.

His previous work has led to publications as first or senior author in multiple well-recognized peer-reviewed journals, including Nature Genetics

and the Journal of Clinical Investigation, and also multiple grants from the Department of Defense, NIH, and the PKD Foundation, and several patents.

This morning, Dr. Fedeles will present on the Advancement of Drug Development Tools for Polycystic Kidney disease As Told Through the PKD Outcomes Consortium Story. So I'll turn this over to Dr. Fedeles. Thank you.

DR. FEDELES: Thanks so much, Christine. Can you hear me?

DR. NGUYEN: Yes, we can hear you.

DR. FEDELES: Perfect.

Presentation - Sorin Fedeles

DR. FEDELES: Good morning, everybody. So, today I will talk about the advancement of drug development tools for PKD as told through the PKD Outcomes Consortium Story. Next slide. So, I'll give a brief C-Path overview because my colleague, Dr. Walls, has done a great job talking about C-Path already, and then I'll talk about PKDOC background and our impact in terms of drug development tool advancement and then finally I'll

talk about our current project under the new iteration that we call PKDOC 2.0. Next slide.

So, C-Path works as a pre-competitive neutral player in the drug development space, and as Ramona has described really well, C-Path brings together stakeholders, including industry, academia, foundations, patient advocacy groups, and regulators and via data and expertise sharing, focused on areas of unmet need, you know, it promotes development of tools that can speed up clinical trials. Our expertise lies at the intersection of data management, curation, biomarker development, disease progression modeling, clinical outcome assessments tool, and regulatory development.

In terms of concentration areas, C-Path is focused on areas that span neuroscience, inflammation, infectious diseases, safety sciences, and rare and orphan diseases. Next slide. In terms of data sets, as a lot of us say, we're only as good as the data that we have, and C-Path has done a great job in accumulating relevant

patient-level data sets, ranging from RCT trials to registries, and as you can appreciate, we've had a great influx of data in the past few years. So, currently, we have more than 450,000 subjects as part of our patient-level databases, with the PKD consortia having quite a large number of data points as well. Next slide.

In terms of the successes, C-Path has been around for I guess 18 years now, and we've had a lot of success in terms of advancing tools and taking them through regulatory endorsement with FDA, EMA, and PMDA. The secret sauce here really is the fact that once these tools are endorsed, once these actionable solutions are endorsed, they can accelerate and de-risk medical product development, and this is key to how we operate and how we impact, at the end of the day, patient health. Next slide.

So, this is the typical structure of our consortia at C-Path. I just wanted to provide a little color. C-Path has an internal team, which is usually an executive director, project manager,

project coordinator, and then we have co-directors that span usually industry, academia, foundations, and then we have industry members and academic members that are part of a certain consortia. And then we create working groups focused on topics of interest usually around regulatory endorsement of tools that address an area of unmet need; so this is a typical structure or consortia at C-Path. Next slide.

This is sort of the microcosm of the greater C-Path slide that I presented. So, we as a consortia, again, act at the intersection of the stakeholders, industry, regulators, academia, and foundations. What really we do is to convene stakeholders and to create and build consensus, and really enable this iterative participation of stakeholders in order to develop methods and develop products that impact the efficacy of drugs. We do this via our neutral convener role in this larger ecosystem. Next slide.

So, ADPKD, as Christine mentioned, is the most common monogenic disorder, kidney disorder

that is. As you can appreciate, the disease is a very slow progressive disease. You have these kidney cysts that form as focal outpouchings derived from kidney tubule cells, and over decades, really, they grow, kidney volume increases, and you have this slow destruction of the healthy kidney tissue while kidney function is maintained via hyperfiltration for a long time, but then a point of no return is reached where you have this precipitous decline in kidney function.

In terms of signs, symptoms, and acute episodes, you have urinary concentrating defects that occur pretty early in the natural course of the disease. Hypertension, again, is an independent risk factor for progression to ESRD in the context of PKD if it occurs before the age of 35, and then you have, obviously, pain due to the mechanical stress, and then acute episodes of cyst rupture, infection, and kidney stones. But as you can appreciate, the disease is very slow progressing, where the functional reserves of the kidney, if you will, are decreasing over time, yet

kidney function remains stable for many years. Next slide please.

So, in terms of the genetics of the disease, as I said, it's the most common hereditary renal disease. It's autosomal dominant while at the cellular level it's recessive. So basically you have germline mutation in either PKD1 or PKD2, which account for the vast majority of cases, and then you have a somatic second hit that triggers this cystic transformation and cyst growth over many years. In terms of prevalence, there are more than 600,000 people in the U.S. and more than 12-and-a-half million worldwide, and there are no common or recurrent mutations. Next slide.

OK, so based on what I've said so far, as you can imagine, when people start thinking about interventions for PKD and possible clinical trial designs, the natural history of the disease works against it. So, you have this very slow progressive disease, heterogeneous presentation, stable kidney function for many years. So in terms of designing trials, this meant potentially long

trials, and based on the hard endpoints, when people start thinking about this, they double of serum creatinine, ESRD, or death, and obviously this makes for a very challenging proposition.

So, very quickly, the unmet need in the field really revolved around finding clinical endpoints or accepted surrogates that can measure disease progression earlier in the course of the disease, where kidney function is largely preserved, and obviously that led to the interest in the development of biomarkers that can be used in drug development, and in particular, biomarkers that can stratify patients into fast or slow progressors; in other words, patients that are more likely to experience progressive disease or not, and also biomarkers that can serve as potential surrogate endpoints for clinical outcomes.

So based on what I told you about increasing kidney size and kidney volume, total kidney volume came into the spotlight very quickly as a very potentially relevant biomarker for PKD. So this is where the genesis of PKDOC came about. PKDOC

started as a collaboration among stakeholders in the field, and the initial mission was to develop a therapeutic area user guide for PKD to develop standard common data elements for PKD, and then work collaboratively to create and integrate a patient-level database from multiple sources in the field, obviously from RCT data, which didn't really exist back then.

And then second best was the registry studies or longitudinal progression studies, and then use those integrated data sets and obviously curate them, map them, and then develop quantitative disease progression models based on those data, and generate consensus in the field regarding the utility of total kidney volume as a biomarker for progression of ADPKD. And finally, because all of these efforts would not be fully impactful without having the regulatory endorsement stamp of approval, obviously the goal was to submit the qualification package of TKV to the regulatory agencies in order to create the maximum impact for stakeholders. Next slide.

So PKDOC started by correlating data sources from academic registries, from the University of Colorado, Mayo, and Emory, in addition to a longitudinal observational study that was sponsored by NIH-NIDDK. As you can appreciate, there were thousands of patients as part of this registry. It was more than 10,000, but as the previous panelists have alluded to, using this type of data has a lot of challenges, and when PKDOC went through the effort of curating and mapping this data, only a subset of patients could be used for this TKV qualification effort. So out of 10,000-plus patients, about only 2300 patients could be used as part of the TKV progression modeling analysis. Next slide, please.

So long story short, after data integration, mapping, modeling, and iterative regulatory interactions, PKDOC was able to qualify total kidney volume as a prognostic enrichment biomarker with FDA and EMA. This is just a diagram that was used as part of the qualification package, and basically, as you can see, irrespective of eGFR

either below 50 mL per minute or above, or age below 40 or above 40, a higher total kidney volume is essentially correlated with a higher probability of a 30 percent decline in eGFR. This is exactly the guidance language that was used as part of this qualification. To paraphrase, the guidance provided qualification for the use of TKV at baseline as a prognostic enrichment biomarker to select patients with ADPKD at high risk of a 30 percent decline in eGFR. So, this was, again, a very impactful outcome of this effort because that meant that TKV could be potentially employed as part of clinical trials to stratify patients. So, next slide, please.

In terms of the enrichment, this is just a snapshot taken from the qualification package, which is partly available. So essentially, when you use a model without TKV versus a model that incorporates TKV, you would essentially require fewer patients to enroll in order to get in one event, in this case achievement of a 30 percent decrease in eGFR. So again, if you extrapolate

this to a large number of patients, this can translate into obviously a significant impact in trial size. So, next slide, please.

As these efforts were ongoing to qualify TKV, again, TKV was deemed as a very useful tool in the field, and it was incorporated in this classification's criteria to stratify patients into classes. This is called the Mayo imaging classification that takes into account TKV plotted versus age. Then based on, essentially, the TKV figures, you can classify patients into classes from 1A to 1E, with 1E being essentially at the highest risk of ESRD. So, this is a useful classification that, again, incorporates TKV as a tool and can be utilized as part of clinical development programs. Next slide please.

The type of mutation actually became an employed criteria to stratify patients as well, so the type of mutation based on PKD1 non-truncating mutations versus truncating, versus PKD2 mutations, led to stratification of patients with essentially the lowest probability of renal survival being seen

in patients that have PKD1 truncating mutations with intermediate probability for PKD1 non-truncating mutations and the highest probability of survival for PKD2 mutations. So again, this became just another criteria to incorporate into patient stratification. Next slide please.

This more recent classification called the propagated score, essentially utilizes the genetic stratification that I mentioned, but also incorporates gender, hypertension events before the age of 35, urologic events before the age of 35, and then essentially leads to a more refined way of stratifying patients based on probability of renal survival. So again, the field has developed a few tools to stratify patients. Obviously, this is very useful and very impactful because this can potentially be employed as part of clinical development programs. Next slide, please.

In terms of our impact, to summarize, PKDOC started as an effort to develop a therapeutic area, as a user guide for PKD, and that led to creation

of our patient-level database, and based on those efforts, those are leveraged to TKV through the successful qualification process as a prognostic enrichment biomarker, and more recently in 2018, TKV was designated as a reasonably likely surrogate endpoint for PKD. So, in theory, it can be used as part of an accelerated approval program to utilize TKV as a primary readout in a phase 3 trial. Obviously, this accelerated approval paradigm requires an acceptable plan for a postmarketing confirmatory trial. Next slide, please.

So the lessons learned as part of these efforts really were that even though TKV had been employed as part of the development programs, the qualification effort quantified the amount of information that was added by essentially using TKV as an original prognostic biomarker. And again, this qualification, per se, has served as a stepping stone to meaningful iterative discussions in the field with regulators about the use of TKV as a reasonably likely surrogate endpoint, and taking it beyond that as a potential surrogate

endpoint for approval.

And the lessons that we learned in terms of using registry data is that, yes, registry data can be critical for establishing the value of a biomarker as a tool in drug development, as we did in TKV, with obviously inherent challenges when it comes to curating the data, to mapping it, and to generating relevant analysis data sets. Next slide, please.

In terms of our current effort under the iteration that we call PKDOC 2.0, -- next slide please -- our efforts are focused on three main areas right now. We are keenly aware of the need to continue data-sharing efforts for PKD and working with our close stakeholders for that. We are very focused on refining the TKV modeling that we had worked on before and developing a clinical trial simulator tool that, again, can be taken through regulatory endorsement and become a stepping stone as part of clinical development programs.

We are also very interested in identifying

novel biomarkers of disease progression or drug response that go beyond TKV, and the third topic is taking a patient-centric approach to both ADPKD and the recessive form of PKD, and essentially generating patient concepts and building PRO tools that can become part of clinical development programs as well. Next slide, please.

So turning to data sharing, I just wanted to stress just how important it is, and it is really the bedrock of everything that we do and that a lot of other organizations do. Why is it important? Data sharing impacts every stakeholder in the field. It impacts academia by improving research, by understanding disease course or variance. It impacts industry by being able to design more effective clinical trials and by understanding and developing biomarkers. And again, at the end of the day, it impacts patients, and this is the most important, and it allows faster drug development. Again, it allows collaborations and allows cross-pollination of ideas in order to drive tools that impact, at the end of the day, patient health.

Next slide, please.

In terms of the modeling clinical trials simulator tool project, again, as I have mentioned already, we have a pretty large patient-level database of registered data, and we have been very keen to acquire other types of data sets, and in particular, RCT data sets. We have acquired HALT, ALADIN, and TAME data sets, and we continue to work with our industry partners to acquire industry-led RCT type data.

What do we do with this data? My colleague, Dr. Walls, has already gone through this in a different context, but we integrate this data, and we use our competencies in data curation and mapping in order to standardize the data and to feed it through our modeling pipeline that is run by our quantitative medicine program.

And, what do we do with it? We build models that are essentially the bedrock of clinical trials simulated tools that can be taken through regulatory endorsement, and that's really the key to success here, going through this entire process,

including the endorsement process, in order to have the most impact for our stakeholders. Next slide, please.

And again, for our CTS model output, the model is intended to be used in clinical trials in order to model disease progression and in order to model trial components or drug effects. At the end of the day, the impact of this tool is really at the level of being able to have a better handle, a more refined handle, on the inclusion/exclusion criteria, enrichment strategies, trial duration, and size, but also this tool can serve as the bedrock of supporting the design of the accelerated approval progress for PKD. Next slide, please.

Again, this is just a snapshot of what a simulator for PKD would look like. This is our Alzheimer's clinical trial simulator tool. I'm just giving you a snapshot. I don't want to comment too much on this, but again, at the end of the day, this tool would be publicly available, and sponsors would be able to utilize that as part of their development programs. Next slide, please.

I just wanted to touch upon PRO-focused approaches because I know Caitlin, the next speaker, will talk about that. Another avenue of high interest to PKDOC right now is to take a patient-focused approach to inform medical product development. As I said, both the dominant and recessive form of PKDs are areas of unmet need and of interest to us, and currently we're using the recessive form of PKD as a case study for organizing an externally-led patient-focused drug development meeting. Next slide, please.

The objectives of this meeting, which is essentially the first step in gathering patient concepts and, down the road, building PRO tools for ARPKD, the objectives are to collect the patient and family experience of living with ARPKD; to get information regarding the factors that influence patients' decision making with regards to entering clinical trials; and also to gather concepts regarding the medical management of ARPKD and the experience that family and caregivers have regarding treatments and aspirations for new

treatments. Again, in terms of the benefit, I don't want to stress that this benefits, obviously, patients, and this is why we are doing it, but also this benefits the entire stakeholder ecosystem, including industry, patient advocacy groups, and, obviously, regulators. Next slide, please.

So, in terms of the value that C-Path and PKDOC brings to the stakeholders, via our drug development tool processes, we can achieve a better understanding of disease and application of biomarkers across stakeholders. We can implement biomarkers in clinical trials, accepted under IND versus qualified, obviously. We can stratify patients, and we can build disease monitoring biomarkers that, obviously, eventually can lead to efficient clinical trials and faster approvals. And most importantly, we can change a patient's journey, and we can take a precision medicine approach to be more successful and more impactful with our drug development programs.

I think that's my last slide, so I want to thank everybody for their attention.

DR. NGUYEN: Great. Thank you so much for an excellent presentation, and we certainly have some questions in our chatbox that we'll try to answer.

I'm very happy to present our second presenter for this session, Dr. Caitlin Nichols. She is the Director of Research at AllStripes Research, a medical data science company with the mission of accelerating new treatments for people impacted by rare disease. In this role, Dr. Nichols oversees scientific communications and the design and execution of real-world data research partnerships with industry, academic, government, and patient advocacy groups stakeholders.

Prior to her current position, Dr. Nichols was a scientific curator on the Product Science Team at 23andMe, where she assisted in the development and improvement of carrier status and genetic health risk reports. She received her PhD in Biological and Biomedical Sciences from Harvard University, where she studied novel cancer

therapeutic approaches, leveraging copy number changes in cell-essential genes. This morning, Dr. Nichols will present on Leveraging Patient Engagement and Real-World Data to Inform Rare Disease Drug Development.

Dr. Nichols, I'll hand this over to you. Thanks.

Presentation - Caitlin Nichols

DR. NICHOLS: Thank you so much for the introduction and to the organizers for the opportunity to speak, and thank you to Sorin for that insightful presentation as well. Today, I'll be discussing some use cases from our work at AllStripes and insights that we've learned about how we can leverage patient engagement and real-world data to inform rare disease drug development. Next slide, please.

Now, all of us here today are familiar with the unfortunate reality that far too few orphan drugs are approved each year. This is despite advances in technology that, in theory, should help to accelerate this space; for example, the decrease

in sequencing costs and improvements in gene editing technology that should expand the field of preclinical programs. However, despite all of the wonderful work that's been done by advocacy organizations, academic investigators, and industry investigators, only 20 rare disease drugs were approved last year. Next slide, please.

We're also familiar with the challenges facing those involved in rare disease drug development. As we know, patient populations are small and geographically distributed, and frequently, these conditions are complex and require care from many different specialties across different institutions. These factors can lead to a scarcity of high-quality data, which can then make it challenging for us to understand how the disease progresses, and impacts both patients and the healthcare system.

Frequently, it's challenging to identify appropriate outcome measures in rare conditions, and this and other reasons can make it very challenging to plan and execute effective clinical

trials for orphaned conditions. Finally, and most critically, forming deep and authentic relationships with patient communities is absolutely critical in rare disease, perhaps more so than in any other indication. Next slide, please.

One tool in our toolkit to address these challenges in rare disease drug development is real-world data, which is data that's collected outside of the confines of the clinical trial. This is what we focus on at AllStripes, and real-world data can help to address challenges in rare drug development across the life cycle.

So beginning in preclinical stages, starting to understand what is the unmet need in this condition; and moving into planning and executing clinical trials, what is the patient journey from when they're diagnosed through to management in the healthcare system, and who are the patients? What are their characteristics at a baseline, and how can we design a clinical protocol that makes sense and is feasible? Then moving into approval and

launch, how is the product being used out in the real world? What's its safety and effectiveness in the real world?

Now, all of these questions are things that can be addressed with real-world data, but -- next slide, -- today I'm just going to focus on the uses of real-world data for the planning and execution of clinical trials. Next slide, please. There are a variety of sources of real-world data, from claims and structured EHR databases, to unstructured clinical notes, and patient-reported or patient-provided data such as surveys or data from wearables. All of these sources of data can be very valuable, but they do have gaps. Next slide, please. And so it's our view that integrating the patient voice is really critical to developing a robust real-world data strategy and filling in these gaps, these four big questions that I'll refer to as the what, who, where, and when, in rare disease drug development. Next slide, please.

So what are these big questions? First of

all, who? Who are these patients? What is the population like at a baseline, and what would be feasible I/E criteria for the trial? Next, what and when? What are the patients experiencing, and at what point in their patient journey? This can help us characterize the unmet need faced by these communities and determine the appropriate outcomes and endpoints that are needed for a trial. And finally, where? Where are the patient's geographically, so we can identify suitable trial sites, but also socially and culturally, so that we can identify appropriate recruitment approaches. Next slide, please.

This is where AllStripes lives, is at the nexus of patient engagement and real-world data generation. Patients and caregivers can sign up to our platform and consent to participate in research in minutes. This research consent is an umbrella consent that allows for the use of de-identified data for minimal risk research, including survey collection, as well as participant recontact over time.

Our team then collects, structures, and analyzes multimodal clinical data from a variety of sources from across the patient journey at no cost to the participants, and then we use the structured and analyzed data to help pharmaceutical companies answer some of these big questions that are potentially blocking their drug development programs. In addition, we provide participants with ongoing research insights and other features to assist them in their rare disease journey.

So today, the case studies that I'm going to discuss are based on our learnings from collecting, analyzing, and working with partners to use these data to help their clinical programs. And while the case studies I'm going to share are anonymized, I'm hopeful that they'll be helpful as you think about your own clinical development programs. Next slide, please.

So the first case study that we'll start with is a question of who, what, and when, and we'll be discussing characterizing the unmet need and the patient journey in a rare pediatric

epilepsy. Next slide, please. In this case study, we worked with a sponsor that was a biopharma company preparing their IND application for a product to treat a rare severe pediatric epilepsy, with seizures beginning in infancy. The challenge in this condition is that there was really a lack of understanding of the natural history and progression of this condition, and in order for the sponsor to better inform their clinical trial design, they needed to better understand the patient journey.

So our solution was to work with this sponsor to develop a natural history study to better understand the needs of the patient community and to help inform their outcome and endpoints selection, and we did this both through participant surveys, as well as through abstracting clinical data from participant medical records.

You can see from the statistics there at the bottom, particularly the bottom-right, over 12,000 individual data points were abstracted for this program for the cohort of less than

40 participants. So it was really a tremendous amount of data characterizing patients with this pediatric epilepsy. Next slide, please.

This slide shows one of the first steps that we do as part of our natural history study development, and this is doing a patient journey map. We create these journey maps by doing a deep comprehensive dive into the medical records of a small number of participants from as far back as their clinical history goes, to birth in this case for the pediatric patient, and looking at all of the different types of clinical documents across the spectrum of care, and we pull out clinical information and information about their journey, really placed in context, so that we can understand not just what was happening but how it related to other events in the patient's journey.

For example, here you can see that we have the birth notes for this patient. They had a normal newborn screen, but shortly after that, they presented to the NICU for seizures, and then they started on their first antiepileptic drug. Shortly

thereafter, they had the first of many AED regimen changes. They were eventually referred to therapies and had genetic testing ordered. Then in blue there, you can see that the causative variant was identified, and they were diagnosed with this rare epilepsy.

We can then track over time additional symptoms as they present, for example, developmental delay, hypotonia, and GI and sleep issues. We can look at assistive devices that patients need, for example, here, a G-tube, monitor; non-pharmacologic interventions, for example, a ketogenic diet. Testing results are shown by the normal EKG and audiology and the abnormal swallow study, and then ultimately we see that this patient was placed on an investigational drug for this condition.

So while this is a zoomed-out view of one of these patient's journey maps for the purposes of protecting participant privacy for this presentation, you can see that this is really a tremendous amount of data in its very deep and

comprehensive way. Of course, this is something that we would love to have for each and every participant in one of our studies, but frequently, due to resourcing, that may not be possible.

So when we work with sponsors, one of the ways that we leverage these journey maps is by, again, doing them on a small number of patients to get this very deep and broad picture of what patients are experiencing, and then we leverage those learnings to carry them into designing our structured data capture for a broad swath of data elements that will be collected from the full cohort; and in that way, we're kind of able to get the best of both worlds.

Now, despite the depth of clinical information here, what's missing is the patient voice and really understanding how the condition impacts participants and their families. Next slide, please. One of the ways that we can address this is through PROs or surveys, and one of the things that we do is surface a survey to every participant on our platform about their symptoms,

when their symptoms first started, what was the first symptom, and what's the symptom that most impacts their quality of life?

In this condition, when we surveyed the participants, we weren't surprised at all to see that the first symptom for the majority of participants was seizures. This is what we would expect. However, when the caregivers were asked about the symptom that most impacted their quality of life, half of them indicated that developmental delays was the most impactful symptom, even more so than seizures, and this is something that we wouldn't have known or necessarily expected without surveying the families. So again, this really underscores the importance of marrying not just the deep clinical data, but also the voice and the experiences of the patients and families to understand the unmet need to be addressed in a future clinical trial. Next slide, please.

This slide shows another example of how we've collected this data in one of our rare conditions. This is dermatomyositis, which is a

rare inflammatory myopathy, and you can see that when we asked these participants about the symptom that most affects their quality of life, the answers were much more heterogeneous than what we saw for the pediatric epilepsy. This survey is something that, as I mentioned, we've surfaced to all participants on our platform that are consented to participate in research. More than 800 across 46 conditions have completed this survey, and this is an effort that we want to continue to deepen and expand on over time. Next slide, please.

Next, we'll move into a case study addressing questions of who, what, and when, and this is characterizing the patient population in a rare metabolic condition. Next slide, please. In this case, the sponsor was a research institution exploring commercialization. They were still in the preclinical stages of development, and they were working on a condition that's a rare inborn error of metabolism. The challenge in this condition is that there's a lack of understanding of how it manifests, including neurological signs

and behavioral symptoms that begin in childhood, and future clinical trials will require appropriate instruments for measuring these symptoms.

To address this problem, we partnered both with the sponsor, as well as with the main advocacy group in this space, to design a natural history study about this condition. The reason why we partnered both with the sponsor, as well as with the advocacy group in the actual design of this study was that it was absolutely critical for us to know what we needed to capture. Because of this condition, and the nature of the signs and symptoms, there's information about the condition that only the families and the caregivers would know when these symptoms are happening, what types of symptoms are happening, so we don't just need their help to collect the data, but even to set the foundation for where we need to start; what's the data that we need to collect? So in partnership with these two stakeholders, we executed clinical data abstraction from participant medical records, as well as surveys. Next slide, please.

This slide gives an overview of how we developed one of the instruments that we used in this study to measure behaviors of the participants. First, the sponsor and the advocacy, KOL from the patient advocacy group, co-developed a comprehensive list of behavioral symptoms and associated data that were of interest for the natural history study. Next, our team developed and tested a survey instrument on our proprietary patient platform with feedback from both the sponsor and the advocate, KOL.

Next, we piloted the instrument to a small group of participants who provided feedback on content, language, and presentation, and then we surfaced the survey to all participants in the study so that they could take the survey if they chose, and we did this longitudinally to track response consistency and disease progression over time. Next slide, please.

This is a high-level overview of the instrument that we developed. The results from this survey are still being analyzed and written

up, so I can't go into too much detail, but just to give you an idea of what we did here, we started with nine different behavior categories -- they're on the left -- and for each of these behavior categories, there were specific behaviors nested underneath them.

If we go to the next slide, we'll see an example. The first behavior category was physical aggression, and there were four specific behaviors we were interested in learning more about: hitting or kicking, scratching, biting, and grabbing. For each of the next behavior categories, there were behaviors nested under them, so a total of 33 specific behaviors that we were interested in learning about in this survey. Next slide, please.

We also had another behaviors category at the bottom, where caregivers could provide free-text information on symptoms that maybe we hadn't thought to include in the survey, and then for each of these behaviors, we asked a variety of questions, for example, about age of onset, triggers of the behavior, and behavior frequency.

Next slide, please.

So what did we learn? We found that, broadly, the results of the survey were consistent with what has been reported in the literature, and we also saw the value of engaging the caregivers in developing this instrument. For example, across three different categories, we found that there were additional behaviors that we hadn't thought to include in the original instrument, for example, one additional physical aggression behavior and a couple of other additional behaviors in these two other categories, Category 4 and Category 8. Next slide, please.

Looking at the other behaviors that were surfaced to us in the free text responses, we found that there was actually an additional behavior category involving eating and feeding behaviors that we hadn't previously thought to include in the behavior survey. In addition, there were at least three behaviors that didn't fit cleanly into an established category, so both of these findings were things that can be carried into future

development of this instrument for potential use in future clinical trials. Next slide, please.

Just returning to the beginning here, speaking about this case study, as you could probably tell from the overview of that survey, it was quite a lengthy survey. The median time to completion was about 20 minutes, and it asked about some challenging issues for the families and caregivers, and yet we had a tremendous amount of engagement on this survey.

For a cohort of less than 30 participants, we collected over 2500 individual survey data points and, really, I think that the reason why the families were so engaged and willing to participate is not just because they understood the importance of this to furthering clinical development for their loved one's condition, but also because we had involved them from the very start, informing the foundation of the study, so they knew that this would be a valuable use of their time because they had been given a voice in what was being collected. In addition, we returned interim results to the

community during the process of survey collection to let them know about what we were finding and help them understand the potential impact of their participation. Next slide, please.

We'll go into a little bit more of a logistical section, in this case, answering who, evaluating I/E criteria for trials, and this insight selection is particularly important because some estimates state that at least a quarter of all rare disease clinical trials fail as a result of challenges with recruitment. So getting these right from the outside is really important as we plan clinical programs. Next slide, please.

For this study, we worked with a biopharmaceutical company that was in the middle of their pivotal trial, and this was in a rare adult onset autoimmune neuropathy, and the challenge here was really recruiting participants for this large multisite trial. So our approach for addressing the sponsor's need was prescreen participants that consented on our platform using data collected from their medical records. We started with

132 consented participants; 112 of those went through the prescreening process, and ultimately fewer than 5 patients ultimately passed the prescreen, and were given the option to be connected to the study site.

Now, you may think, well, that's a pretty small number. Why are you using this as a case study about I/E criteria? Next slide, please. And really, I share this to underscore the importance of thinking carefully about I/E criteria, particularly in rare disease clinical trials, so I'll share first the top medical reasons that patients failed the prescreen.

The first was for a diagnosis of diabetes. This is, of course, a common condition; 1-in-10 Americans have a diagnosis of diabetes. This is a population typically of middle-aged to older adults, so already a higher likelihood of having a diabetes diagnosis, but diabetes is also a known comorbidity in this condition with 15 to 20 percent of individuals living with this condition also having a diabetes diagnosis. So,

ultimately, 9 patients were screened out initially because of a diabetes diagnosis. Next slide, please.

The second most common medical reason for failing prescreening was a history of malignancy; again, a very common diagnosis; 1-in-2 people in the U.S. will have a cancer diagnosis over their lifetime but, again, as this is a population of middle-aged and older adults, they're more likely than the general population to have had a cancer diagnosis at some point in their medical history.

So these numbers, since these are only 17 patients, this may look relatively small compared to the 112 that were screened, but the point that I'd like to make here is that these are just a subset of the exclusion criteria for this trial. This trial had at least 10 different exclusion criteria, each of which resulted in patients being screened out, and I haven't included the smaller numbers, again, in the interest of protecting participant privacy.

But when we're working in rare disease,

every potential participant counts, so this is not to say that these I/E criteria were inappropriate for the condition. They may well have been appropriate and should have been included, but this is just the importance of really thinking carefully about the characteristics of the population and whether these I/E criteria are going to be feasible, based on the sample size that you need and the underlying characteristics of the population that you're working with. Next slide, please.

Next, we'll move into a discussion of the where, identifying appropriate trial sites. Next slide, please.

Zooming out, we surfaced across all participants on our platform a survey about past clinical trial participation, as well as interest in participating in a future clinical trial, and more than 450 participants took this survey. We found that nearly three-quarters of participants have not yet been involved in a clinical trial, but about three-quarters of participants are either extremely

or very interested in participating in a future clinical trial. Next slide, please. But when we asked participants what would be the biggest barriers to them participating in a trial, the most common answer was distance to a potential study site.

While we knew that this is something that would be a barrier to participants, I was surprised that this was the most common answer, even above potential risks of study participation or negative side effects of the experimental treatment. So this really underscores the importance to potential participants of this travel burden piece of enrolling in a trial. Next slide, please.

The magnitude of this burden is underscored by some analyses that we did of participants on our platform. This is across 900 participants in 36 different conditions, and this was done in the fall of 2021. We took each participant and determined the distance from their resident zip code to the nearest trial site in their condition, both for interventional and

observational trials, and then we bucketed the patients into condition categories, and took the median distance among those patients, and that's what you see here in the graph.

For example, if we look at the other systemic category in the navy blue, the middle-of-the-pack patient would have to travel more than 800 miles to get to the nearest interventional trial site in their condition, a tremendous distance. And even in the conditions with the lowest travel burden, here, for example, tumor and lymphatic conditions and epilepsy conditions, the middle-of-the-pack patient would still have to travel more than 100 miles to get to the nearest trial site. So what are some ways that we can address this using real-world data? Next slide, please.

I'm going to return for a moment back to the case study of the company that was recruiting for the pivotal trial in the adult onset autoimmune neuropathy, and as I mentioned, fewer than 5 patients passed the prescreen and were given the

option to be forwarded to a clinical site. And what I didn't mention is that 40 participants actually dropped out of the prescreen because they were too far from any of the trial sites that the sponsor had set up, so that was a major barrier to patients participating. But there are some ways that we can think about addressing this if we go to the next slide.

Starting in November of 2021, we had 71 participants in this condition on the platform. At the time, there were six different clinical trials in this condition, with 15 trial sites across all of them. So looking here, we can see that more than 55 percent of participants at the time lived at least 200 miles from the nearest trial site. We wanted to try to find patients that were less than 200 miles from a trial site, and we were involved in targeted recruitment within 200 miles of those trial sites for the trial that we were helping to recruit for. During that time as well, 10 trial sites were added across all of the six different trials that were ongoing. Next

slide, please.

By February of 2021, we had 111 consented patients on the platform. There were still 6 trials happening in this condition, but there were now 25 sites spread across those six different trials. And as a result of our targeted recruitment, as well as the addition of these trial sites, we saw that about two-thirds of participants were now less than 200 miles from the trial site. And if we go to the next slide, this is a 57 percent increase right where we want the patients to be, either a short or an intermediate distance from the trial site.

So while many of the sites had been established prior to us becoming involved in this project, there are ways to address challenges, for example, by engaging in targeted recruitment once you've actually selected the sites. But are there ways that we can better inform site selection ahead of time to kind of get around some of these issues?

If we go to the next slide, we see an example in our lysosomal storage disorder cohort.

This is an analysis that we did across 9 lysosomal storage disorders, 151 participants in total, and this heat map shows the geographic distribution of these patients by U.S. census divisions. Next slide, please. We wanted to identify prospective centers of excellence for lysosomal storage disorders. Some conditions do have them, but others don't under this umbrella of LSDs, so to do this, we evaluated care centers based on four different criteria.

First, did they have multidisciplinary care teams? Next, had they participated in at least one peer-reviewed publication in an LSD in the past? Third, had they hosted a clinical trial in the past in a lysosomal storage disorder? And then finally, did they have a metabolic genetics clinic?

When we performed this analysis, we found 54 centers met all four of these criteria, and 22 centers met three of the criteria. We can notice, in particular, that there's a relative dearth of these prospective centers of excellence in the Rocky Mountain region, as well as in the

Upper Midwest, and portions of the Southeast United States. Next slide, please.

Just how far would participants on our platform have to travel to get to one of these prospective centers of excellence, either for care or for participating in a clinical trial? Here we took, again, the distance from where the patient resides to the nearest prospective center of excellence, and then this box plot shows the distribution of those values.

Here, the middle-of-the-pack patient would have to travel almost an hour and 45 minutes to get to the nearest center of excellence and nearly 100 miles. This represents more than 4 times the travel time that the average American had to travel for healthcare in the year 2000 and more than 9 times the travel distance that the average American had to travel for healthcare in the year 2000; again, just to underscore how potentially burdensome this is, even with going just to the nearest center that we've identified. Next slide, please.

We also wanted to understand if this travel burden varied by region of the country. As I mentioned, we saw that there were some pockets of the country that seemed to lack prospective centers of excellence, so we did the same analysis where we found the shortest distance from the patient to a center of excellence, and then bucketed the patients by region, and we showed here the median distance of the patients in each region.

We can see that patients in the west-north central -- that's the upper dark teal Midwest region -- the east-south central -- that's the purple region there in the southeast United States -- and the mountain region -- that sort of median teal color -- have the highest travel burden.

So while this may not be too surprising based on population distribution and geography, the LSD patients on our platform that live in those three regions account for nearly a quarter of all participants in our cohort. And as mentioned previously, every single participant counts, so

what are ways that we could potentially address this?

Aside from identifying these centers of excellence, perhaps it's worth looking into if there are other care settings where trials could be administered for these patients. For example, are their community settings that would be equipped to host a trial site to help diminish some of this travel burden? Next slide, please.

Another potential solution is brought to light by an analysis that we performed, a survey where we asked patients about their use of telehealth and their attitudes toward telehealth during the COVID-19 pandemic. More than 700 patients on the platform responded to this survey, and 78 of those participants had an option to participate in telehealth and had used telehealth on at least one occasion.

Of those 700 participants, 74 percent of participants indicated a preference for telehealth, either whenever possible or at least for some types of appointments. And while telehealth and virtual

trials aren't necessarily a one-to-one, we believe that this may indicate an openness on behalf of rare disease patients to participate in sightless trial models. There are many organizations that are innovating in this space, and we encourage that continued innovation.

Obviously, one solution, virtual trials or targeted recruitment, is not going to be the cure-all for the challenges of diminishing travel burden for participants but, again, just being aware of these different options and the importance of making trials feasible for patients, and not just that they can be enrolled, but also to diminish the number of patients that are lost to follow-up over time. Next slide, please.

If there are a couple of things that I would want you to take away from the presentation today, the first is that real-world data can help to address the challenges that are inherent in orphan drug development, and there are gaps in real-world data, and it's our view that integrating the patient voice is critical to answering these big

questions in drug development, particularly when it comes to trial planning.

Who are the patients at a baseline? What's the characteristics of the population? What are they experiencing, and when? What are the most impactful outcomes for us to address with a trial? And finally, where are the patients? How can we make trial sites and recruit patients in a way that makes sense so we can meet these recruitment goals? Next slide, please.

I'd like to end, of course, with the reason why we do what we do, which is the patients and families impacted by rare disease. It's our mission at AllStripes to accelerate treatments for these folks, and moving forward from these case studies that I've shared, we are going to double down on how to further incorporate the voice of the patients to empower them, to provide data that can then help to accelerate treatment for their diseases and the diseases of their loved ones.

Thank you so much for your time and attention, and I'm happy to take questions, and I'm

looking forward to the discussion.

Q&A

DR. NGUYEN: Thank you so much, Caitlin. That was outstanding; such great information.

I thank Drs. Fedeles and Nichols for sharing their impactful insights on how to optimally leverage data collected from rare disease patients to inform drug development in that space.

At this time, we'll transition to the panel discussion, where it's going to be a short panel discussion because we're running out of time. I just want to briefly introduce Dr. Aliza Thompson, who is the Deputy Director of the Division of Cardiology and Nephrology at the FDA, that oversees therapeutic development for the treatment of cardiovascular and kidney disease. She has been with the agency since 2007 and has been widely recognized for her significant contribution to public policies to improve outcomes for patients with renal disease.

Thanks for joining us, Aliza. It's great to have you.

I am actually going to start off with a question that we obtained prior to the meeting. It is, what are the tasks that patient advocates should undertake in order to accelerate the process from research in the lab to trials?

Sorin, do you want to go ahead and take a stab at that?

DR. FEDELES: Sure. Can you hear me, Christine?

DR. NGUYEN: Yes.

DR. FEDELES: Again, patient advocacy groups are partners. They can work with basic translational clinical scientists to create opportunities to connect a dispersed patient population to research, as we heard from Caitlin, to encourage research funding, to shape proposals, and to really, at the end of the day, help design clinical trial protocols.

At the end of the day, it's about connectivity, it's about collaboration, and really engaging patient advocacy groups as key stakeholders as part of this ecosystem can result

in a better understanding of indications, like PKD, for example, to identify targeted therapies and refined standard-of-care therapies.

I think they're a key partner as part of this process, and there's no magic bullet, and there's no recipe for how to do it exactly. It's about connectivity and collaboration, data sharing, and staying at the forefront of pushing efforts forward. That's what I would say from our experience in the PKD space.

DR. NGUYEN: Great. Thank you so much.

DR. NICHOLS: Sorry. I'll jump in if that's ok, Christine.

DR. NGUYEN: Absolutely. Please do. Thank you.

DR. NICHOLS: I couldn't agree more with what Sorin said and the importance of patient communities as partners. I think my advice for advocacy organizations would be to get involved with the investigators as early as possible in drug development, even if it's the folks who are working in cells or mouse models. I don't think it's ever

too early to begin to share what's really impactful, and make sure that from the start, those relationships and that knowledge is being shared so that they're focusing on what's most important to the patients and not aiming for some outcome that isn't going to ultimately improve patients' quality of life.

The other thing that I would say is the importance of patients in helping to educate each other on the importance of different research opportunities or about different trials. I think we try to provide lay friendly and public friendly accessible research, but I really think it's so impactful when it comes from the community itself and folks you can speak to, as this was my experience, this is my advice for participating, or not participating, or what-have-you.

Just really having that voice and speaking with the community can be so impactful to help galvanize others and help them understand the importance of your research efforts.

DR. THOMPSON: Maybe I'll jump in, too, for

this one, because I think it is just a fabulous question. Obviously, successful drug development takes an understanding of mechanism and basic science, and a lot of basic science research. But, really, to make that translation to enable successful drug development, people need the toolkit. Sponsors need the toolkit to actually do trials in an efficient manner and effective manner. So I think patient advocacy groups really play a critical role in making sure they have that toolkit. They can have the biomarkers and the tools they need to understand the patients who are likely to progress and have a way to help measure response, and potentially surrogate endpoints. I think that comes from helping with some of these studies that are done, but also really advocating for data sharing.

DR. NGUYEN: Thank you.

I'll just chime in that the rare disease space is where we don't have the option to be inefficient; that's the bottom line. We're rushed for time, right, because it's a great area of unmet

need, and we have a very limited number of patients. So I think it's critical that there is as tight of a collaboration between patients, their families, advocacy groups, sponsors, and working with the FDA.

Ultimately, we're in charge of making sure that the drugs we approve are safe and effective for our patients, and there's science and there are regulations to support that. And the sooner all of our messaging gets together, everyone understands each other's perspective and what the needs are. I see really a big collaboration that needs to dance well together, and having a one-piece silo and another really introduces inefficiency that we can't afford.

So I think that's the overarching message, and certainly for us working in FDA, that's the vision we hope that everyone will buy into because at the end of the day, that's what's going to give our patients and families what they need, and ultimately that's who we serve.

I wish we had another 20 minutes to our

panel discussion, but I'm mindful of the time. So at this time, I want to thank Caitlin, Sorin, and Aliza for helping us with Session 2, and at this time, I will turn the meeting over to Kerry Jo for concluding remarks for our day 1. Thank you very much.

Concluding Remarks - Kerry Jo Lee

DR. LEE: Hello, everyone, and welcome to the end of day 1. I really want to thank everyone who participated in today's incredible session, the moderators, all of the speakers, as well as the behind-the-scene staff such as Audrey Thomas from the Rare Diseases Team and Jill Curran from Johns Hopkins, and the AV team to make this happen.

We had close to 2,000 registrants for this workshop, and many of you sent questions in advance, which were really helpful to inform our discussion, both for today's presentations but also future engagement. If we know what it is you want to learn about, it's helpful for us to construct future sessions that will be informative. For anyone who missed it or would still like to view,

or review, the workshop, given the tremendous amount of information and resources that our speakers provided, our intention is for these to be accessible online in perpetuity, either from the FDA-CDER ARC webpage, as well as Johns Hopkins CERSI webpage.

A few take-home points I think we heard today in Session 1 on how to collect quality and fit-for-purpose data, the FDA does have a real-world data, real-world evidence hub resource online that has a tremendous amount of information, including demonstration projects and key guidances that are critically important. So please seek that out for our latest thinking on how to use the real-world data inthe development of real-world evidence.

The power of integration of data; rare diseases are rare, and having data silos is really not helpful and creates additional challenges and can impede rare disease drug development; however, there are a lot of considerations to keep in mind when you're trying to integrate multiple data

sources to ensure that they're fit for purpose and informative. There are resources and tools available to stakeholders to help with these considerations, and there are fundamental principles of data sharing that really needs to be thought about early, such as consent, as well as what your data standardization is going to be and the data model that you're going to follow to inform you how to optimally collect data based on your setting, such as perhaps a clinic or your goals.

In Session 2, we really learned a lot about the uses of data sources to inform rare disease drug development. We learned that learnings from this data can be used to support qualifications and fit-for-purpose tools for use in rare disease drug development trials, but also that we need to be as thoughtful as possible about trial design. Data is critical to supporting the translational strength to support potential biomarkers as surrogate endpoints for direct clinical benefit, as well as the utility of other novel endpoints, the selection

of the right trial population, enrichment strategies, and other aspects of trial design.

Today we focused also on the real-world data use for the planning and execution of these clinical trials and that there are potential roles for the use of real-world data across all phases of drug development. We hope the case studies were particularly informative, and as well, a really critical point is that the patient experience is critical to defining the unmet need to be addressed in clinical trials, as well as designing the optimal trial for the patient population to be enrolled in, considering the overall logistics of conducting a trial. And in the end, it takes all of us to advance rare disease drug development.

So if I could just leave you with one final thought, I would say, after our session today, it would really be, when we embark on the collection and use of real-world data and real-world evidence, you need to start with the end in mind. If you're looking to inform elements of future or current trial design, you have to ensure you're collecting

the right elements.

So the right elements, at the right time intervals, and the right patient population, you have to be thoughtful about data collection, standardization, and models to ensure that what you're collecting will be fit for use. And when it comes to the use of what you've collected, it is not one-size-fits-all. There are unique aspects of individual rare diseases and potential therapies that will affect how you can use the data you've collected. There are factors of the condition, the physiology of a disease, predictability of the natural history, characterization of natural history, and there are also factors to consider when it comes to the design of the clinical trial in which the data is going to be utilized. So the endpoint selection, the subjectiveness, or objectiveness, are relevant to the endpoint in the population studied, as well as the effect of potential therapy, whether that's modest or large effect.

So today's been a really important

discussion in the collection and use of fit-for-purpose data for rare disease drug development. We hope to see you all tomorrow as we move forward into a discussion on how to use data from small populations and how we can approach and think about the design and analysis methods, another big challenge for clinical trials in rare diseases. Thank you all so much.

(Whereupon, at 12:02 p.m., the workshop was adjourned.)