UNITED STATES OF AMERICA

DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

+ + +

CENTER FOR DEVICES AND RADIOLOGICAL HEALTH

MEDICAL DEVICES ADVISORY COMMITTEE

+ + +

VIRTUAL PUBLIC WORKSHOP - SPINAL DEVICE CLINICAL REVIEW

+ + +

September 17, 2021
8:00 a.m.

Via Zoom Videoconference

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

WORKSHOP PARTICIPANTS:

CAPT RAQUEL PEAT, Ph.D., M.P.H., USPHS
Director, Office of Health Technology 6 (OHT 6: Orthopedic Devices)
Office of Product Evaluation and Quality
CDRH/FDA

CHRISTOPHER D. HARNER, M.D., FAOA
Deputy Director, Office of Health Technology 6 (OHT 6: Orthopedic Devices)
Office of Product Evaluation and Quality
CDRH/FDA

HASSAN SERHAN, Ph.D.
Prestige Adjunct Professor
Bioengineering Department
University of Toledo

JUSTIN EGGLETON
Vice President, Spine Regulatory Affairs
MCRA, LLC

VINCENT J. DEVLIN, M.D.
Chief Medical Officer
Office of Health Technology 6 (OHT 6: Orthopedic Devices)
Office of Product Evaluation and Quality
CDRH/FDA

SYDNEY GIBSON, Ph.D. (on behalf of Elizabeth Adegboyega-Panox, M.D.)
Division of Health Technology 6 B (Spinal Devices)
Office of Health Technology 6 (OHT 6: Orthopedic Devices)
Office of Product Evaluation and Quality
CDRH/FDA

HONGYING JIANG, Ph.D., RAC
Safety Signal Manager
Office of Health Technology 6 (OHT 6: Orthopedic Devices)
Office of Product Evaluation and Quality
CDRH/FDA

CAROLINE MOAZZAM, M.D.
Medical Officer
Office of Health Technology 6 (OHT 6: Orthopedic Devices)
Office of Product Evaluation and Quality
CDRH/FDA

DIRK ALANDER, M.D.
Medical Officer
Office of Health Technology 6 (OHT 6: Orthopedic Devices)
Office of Product Evaluation and Quality
CDRH/FDA

PAUL ANDERSON, M.D.
Professor of Orthopedics
Wisconsin School of Medicine and Public Health

DARREL BRODKE, M.D.
Jack and Hazel Robertson Presidential Endowed Professor
Chair, Department of Orthopaedics
University of Utah

S. RAYMOND GOLISH, M.D., Ph.D., M.B.A.
Chief Medical Officer
HCA Healthcare JFK Hospital

KHALED KEBAISH, M.D.
Division Chief, Orthopaedic Spine Surgery
Professor of Orthopaedic Surgery and Neurosurgery
Johns Hopkins University School of Medicine

A. NOELLE LARSON, M.D.
Associate Professor of Orthopaedics
Director of Research, Division of Pediatric Orthopedics and Scoliosis
Mayo Clinic (Rochester)

JANICE HOGAN, J.D.
Partner
Hogan Lovells U.S., LLP

RONALD JEAN, Ph.D.
Director, Division of Health Technology 6 B (Spinal Devices)
Office of Health Technology 6 (OHT 6: Orthopedic Devices)
Office of Product Evaluation and Quality
CDRH/FDA

COLIN O'NEILL, M.B.E.
Assistant Director, Extracolumnar Spinal Devices Team
Division of Health Technology 6 B (Spinal Devices)
Office of Health Technology 6 (OHT 6: Orthopedic Devices)
Office of Product Evaluation and Quality
CDRH/FDA

BARTON SACHS, M.D., M.B.A.
Medical Officer
Office of Health Technology 6 (OHT 6: Orthopedic Devices)
Office of Product Evaluation and Quality
CDRH/FDA

JIANXIONG CHU, Ph.D.
Acting Director, Division of Clinical Evidence and Analysis 2 (Biostatistics)
Office of Clinical Evidence and Analysis
CDRH/FDA

JOHN C. BENSON, M.D.
Neurologist
Mayo Clinic (Rochester)

INDEX

INDEX

M E E T I N G

(9:00 a.m.)

CAPT. PEAT: Good morning, everyone, and welcome to the Spinal Device Clinical Review Workshop. My name is Captain Raquel Peat, Director of the Office of Health Technology 6, which is in the Office of Orthopedic Devices in the Center for Devices and Radiological Health. Today is a busy day, as I have dual opening remark responsibilities for meetings held by FDA, as well as the United States Public Health Service, hence the reason I am in my operational dress uniform, which is the uniform of the day during this pandemic period. As such, my remarks will be brief.

To start, we really wish that this workshop could be held in person; however, we are still in the midst of the COVID-19 pandemic which continues to affect all of us in so many different ways. FDA staff continues to be dedicated to their work to bring diagnostic devices, personal protective equipment, and therapeutics to the American public as expeditiously as possible.

I would like to recognize the selfless individuals across the Agency and in particular, the staff in the Office of Orthopedic Devices who have continued to assist with COVID work beyond their duties as it relates to the regulation of orthopedic devices.

Despite it all, having this meeting during this challenging period in our history should underscore the importance that we place on this effort and this community. I'm really thankful for your interest in this event and I'm thrilled that so many of you could join us virtually.

Before getting into the details of today's events, I would like to remind everyone of FDA's mission to promote and to protect the public health. The Center for Devices and Radiological Health achieves this by ensuring patients in the U.S. have access to high-quality, safe and effective medical devices of public health importance first in the world.

Patients may include your family, your friends, even yourself, and are the primary focus of our mission each and every day. In focusing primarily on our patients, OHT 6 has adopted a vision, which is to be the world leader in the science and regulation of orthopedic medical devices. In pursuing both the vision of the Center and the Office of Orthopedic Devices, we are systematically ensuring that we are providing an atmosphere for managers and staff to lead in a supportive and creative work environment while implementing changes that positively affect those within the orthopedic ecosystem. Overarchingly, this effort and other similar activities support CDRH's vision.

To achieve the mission of FDA, through the various visions for those within the orthopedic product sector, OHT 6 has committed to sharing knowledge with external stakeholders to enhance public understanding related to medical device regulation and to improve the knowledge of every person involved in the orthopedic medical device life cycle.

While this public workshop is not intended to communicate any new policies, processes or interpretation regarding medical device marketing authorization, our hope is that it will serve to facilitate further efforts, discussions, and collaborations in the regulation of spinal devices.

Just over 1 year ago we held the Center's first virtual workshop in the height of the COVID-19 pandemic and have completed a number of virtual workshops in 2020 and earlier in the year. Today's meeting represents the fourth virtual workshop from the Office of Orthopedic Devices for 2021 and we have two more events planned to close out the calendar year.

Today's workshop will focus on spinal device clinical review. Clinical trials represent the pinnacle of the performance data supplied to support a medical device and requires much time and cost to plan, run, analyze after results have been collected and presented within a regulatory submission. Please note that while our regulations distinctly

differentiate spinal, neurological, and radiological devices, we recognize that orthopedic spine surgeons, neurosurgeons, and interventional radiologists are among the specialties that use products that are designated as spinal devices by FDA.

Today we embark on a journey to begin a conversation on spinal device clinical review. Given that we're limited to this 1-day meeting, we have carefully selected topics for enhanced discussion. We'll kick off the meeting with a presentation on nonclinical performance testing and the role and values of these bench tests, followed by commentary from a member of the medical device industry. We will then provide an overview of spinal device clinical review with a focus on types of clinical evidence, patient assessments, and evaluation of device safety and effectiveness. This presentation will be followed by a talk on strategies for enhancement of diversity in spinal device clinical trials, and we will also share an internal analysis of sex, race, and ethnicity data from prior spinal device clinical trials.

We are honored to have key opinion leaders and experts from the spinal surgical community who will then discuss how to define a target population, enrollment criteria, and provide additional dialogue and strategies for inclusion of underrepresented groups. This session will be followed by commentary from a member of the medical device industry on clinical trials and a question and answer session.

After we return from the lunch break, we will have a brief survey which everyone is encouraged to participate in, and for the afternoon we will have two more clinical discussion sessions and Q&A sessions related to the topic of evaluation of safety and effectiveness for spinal devices. The last clinical discussion on spinal device effectiveness will include a guest radiologist who, together with our honored spine surgeons, will delve into the topic of spinal fusion through a discussion on case presentation.

The day will be quite interactive. We welcome your questions or comments. Please

participate in the Mentimeter survey as your viewpoints matter and will provide insight for additional next steps as we further regulate orthopedic medical devices.

This workshop could not be accomplished without the multiple individuals at the helm. I wish to thank in advance all of today's moderators and speakers, specifically. I would also like to thank all of the members of the Office of Orthopedic Devices, the Office of Clinical Evidence and Analysis, the Office of Communication and Education, to include FDA Studios, who have contributed to the contents and planning of today's workshop and equally, I would especially like to thank our external participants who are generously providing their time and expertise today. To everyone else, thank you for your participation as you are an instrumental member of the orthopedic community.

Lastly, I would like to introduce Dr. Christopher Harner, who will be the Master of Ceremonies for today's event. Dr. Harner recently joined the Office of Orthopedic Devices as part of my senior leadership team as a clinical deputy office director and is dual certified in both orthopedic surgery and orthopedic sports medicine. He spent the past 35 years as an academic orthopedist both at the University of Pittsburgh and the University of Texas. He brings a wealth of orthopedic experience and clinical leadership to our organization. Thank you, Dr. Harner, for your willingness in your short tenure with the Agency to serve as the emcee for this workshop. I will now turn the meeting over to Dr. Harner. Thank you all for your time and I hope that you enjoy the rest of the meeting.

DR. HARNER: Good morning, and thank you, Captain Peat, for that kind introduction. It is my pleasure to be a part of the OHT 6 team and I am honored to be the Master of Ceremonies for today's workshop.

We have an excellent program today and with presentations from the FDA staff, as well as special guests from the clinical community and medical device community. Our focus today, of today's workshop, will be on the FDA clinical review of spinal devices. A

detailed agenda and full biography of each speaker is available on the FDA website.

Now, for a few logistics. Our designated staff will monitor your comments and questions throughout. Please send your comments to the following e-mail: OHT6-Feedback@fda.hhs.gov. Any remaining unanswered questions or comments will be saved and addressed subsequently and, as appropriate, incorporated into the conference proceedings.

Please note, upon our return from the lunch break there will be an audience survey that you will be asked to complete, we look forward to your valuable input.

With these housekeeping matters taken care of, we will now proceed with today's program. Can we have Dr. Serhan's first slide, please? It is my honor to introduce our first session speaker, Dr. Hassan Serhan, who is the cofounder and treasurer of the International Musculoskeletal Society and a Prestige Adjunct Professor in the Bioengineering Department at the University of Toledo. He is an influential leader in spine, clinical biomechanics with deep expertise in spine and orthopedic technology, evaluation, and development.

Dr. Serhan.

DR. SERHAN: Hello, everyone. I would like to thank Dr. Ronald Jean and Dr. Vince Devlin for this kind invitation to talk about the roles and limits of nonclinical performance testing of spinal implants.

The presentation objectives: First, we're going to talk about the clinical and biomechanical goals of these spinal implants, then the importance of biomechanical cadaveric and animal testing, the role of FDA, as well as when we need to do clinical evaluation.

When we think about spinal implants, we think about two different groups. One are the fusion devices, the other one is the motion preservation devices. The fusion devices, they are more of an internal brace until fusion takes place. Motion preservation devices,

they last the life of the patient. So, with fusion devices we like to provide stability across the operative segment until fusion matures and takes over, as well as we want the implant to reduce the graft-related failures and complications, like graft collapse and subsidence, maintenance of the sagittal balance, and prevent post-operative kyphosis, the non-fusion devices, all of the above plus, lasting the life of the patient.

Clinical performance of spinal implants is impacted by several factors, one of which we discuss a lot: the design, materials. However, the manufacturing process, post-manufacturing process, as well as the intraoperative manipulation play a major role in the performance of these devices.

So why do we do test these implants? As engineers, we think about spinal implants as durability, strength stiffness, stress shielding, subsidence, migration, corrosion, wear, imaging. From a clinical point of view, you look at them and they are more of a clinical failure. When you have an implant failure, you need to revise it, so those are revision, loss of correction, pseudarthrosis, nerve compression, nerve impingement, late infection due to wear debris, and diagnostics, if you have imaging ablation. These are different failure modes of implants.

Up left here, you can see a screw, typical screw failure. This is a dual rod failure, another rod failure. Here, a set screw loosening and a rod failure, and the rod migrated to the buttocks of the patient. These are collapse of an expandable cage and a cage migration that necessitated a revision here. This is a cage subsidence, you can see the cage embedded inside the vertebral body totally. This is a cage migration. This is a pseudarthrosis due maybe to stress shielding. Here is a fractured polymeric cage. We've seen more fractures even with a 3-D titanium printed cage.

These are complications associated with total disc replacement, osteolysis due to excessive wear debris, you could see the pockets here, a total disc dislodgement and failure.

This is another dislodgement of the core and a failure. Total subsidence, you see bone on bone here. This is a total device migration, the cervical spine. Another key disc migration. I'm sorry, disc migrations. You would see a lot of those subsidence and migration.

When I talk about the manufacturing process, it needs to be documented. Why? If you're machining your part, the first part you machine and the second one you do electrical discharge machining, EDM, you'll get the exact geometry. However, your fatigue performance will be reduced for Ti6-4 from 73 to 24. Therefore, when you get a new vendor you need to specify the manufacturing process to be used in the drawing and the production.

Laser etching, the effect of laser etching on fatigue performance, well documented. Now there's a modification of this laser etching to reduce that impact. You could see here almost 50% reduction due to laser etching. Having the laser etching at the end of the rod maybe could solve that issue.

Intraoperative techniques, they affect the fatigue performance, they affect failures of the implant set screws and, as you can see from these images here, if you had them on axial screw and you have the rod, you could torque the screw, set screw to the limit, but the rod is naturally seated. When the patients start moving, the rod will loosen and you could see a migration. So maybe a revisiting of the set screws could do that.

Another thing, in a lot of the fatigue testing, the ASTM standards, we don't request the set screw loosening values, which it should be monitored.

Talking about the ASTM standard, we mimic the corpectomy model or a compromised anterior column to mimic the failures and you could see here screw failures due to bending with this one. That thick, you'll find bending. This is a straight rod, the brown one. When you bend the rod at different techniques, now you're reducing the fatigue performance almost by 50%. A bent rod that is stress relieved, it goes up. This is for

Ti6-4, this study we've done in the past.

When we do the ASTM testing, F1717, we test a straight rod, but in a patient you may kink the rod that much when you're doing osteotomy and you could see the result of kinking that rod, the bilateral failures. This is a study showing actually the effect of using a satellite rod or railroad rod like double rod on one side here and using a satellite rod. This is by Dr. Gupta, he's done that study.

Loss of correction, it could be due to two reason, the stiffness of the rod, which you get by standard testing, or polyaxial screw loosening, which it could be a fatigue issue. So the post-op correction loss could be immediate or it could be over time when the set screw loosens and the polyaxial screw failed.

When we talk about testing of spinal implants, the interbody devices, we do a compressive shear to mimic the shear force at the L5/S1. We do a push-out/pull-out of the devices; however, we don't do a lot of impact testing. I think there is an issue now to create that standard. Subsidence, we discussed this one also before.

In terms of motion preservation, we correct -- statically, but we do fatigue testing under complex loading to mimic the in vivo motion and to create wear, we use a modified device tester for that.

In terms of the cadaveric studies, they are very important to show us how the device is going to function in vivo, so the functionality of the device is evaluated in cadaveric study, as well as the evaluation of the surgical technique to implant that device. Do we damage the endplate, do we destabilize the spine during the procedure and evaluate the bone implant interface as well with these cadaveric studies.

When we think about FEA, a lot of the times we use it for determining the worst-case scenario. But FEA is very valuable, actually, to evaluate the effect of instrumenting a segment and how's the impact on the adjacent level as well as the bone implanted interface

stresses. So it is under-utilized today because you could simulate not only the ligaments, but the muscle. So a validated FEA could be very, very valuable. In this instance here we are looking at the effect of using PEEK rods instead of titanium, and what's the impact on the fusion in the adjacent level.

Imaging. Here's a titanium threaded cage, you cannot see the bone, if it's fused or not, with the standard X-ray. Here you could see the incorporation of the bone with the endplate with a standard bone grafting.

MRIs, imaging. Again, this is stainless steel. If you're using stainless steel implant you may not be able to see the neural elements. I mean, here you could see this titanium screw and you could see the -- very clearly in these images.

When there is a new technology or a novel technological feature presenting additional or unknown risks, we may need additional preclinical studies, but when this additional nonclinical performance is not adequate, we will need to go and perform a clinical evaluation.

So you could see here with motion preservation, we do the fatigue testing to look at the wear, we put it in cadaveric studies trying to look at the motion and the level as well as adjacent level and then back on the entire spine.

The wear debris generated here we injected in a rabbit, trying to look at the effect of the wear debris on the neural elements and after that, implanted in a goat neck, for example, if you're doing cervical spine or a lumbar baboon or a sheep or a pig before we go to the clinic. So a lot of these preclinical studies are required before you go to the clinic and if they are not adequate, you must go to the clinic.

So a validated FEA may be the easiest way to determine your worst-case scenario. The worst-case scenario biomechanically may not be the worst-case scenario clinically. A largest stiffest cage is a worst-case scenario from a stress shielding. The smallest case, the

smallest cage is a worst-case scenario from a mechanical point of view. So if you're using the largest stiffest cage, you may stress shield the bone inside the cage while if you're using the smallest cage, it may telescope into the endplate, so you need just to understand what is the worst-case scenario.

The effect of intraoperative manipulation needs to be evaluated and it needs to be considered. For example, the expandable cage, the testing we do. Do we do it at a full expansion, a half expansion, a quarter expansion, because that may change the fatigue performance or the collapse of these devices. Remember that a change in the manufacturing process may totally change the fatigue performance of your devices, so it needs to be quantified.

Why do we use the ASTM standards? Standard property stiffness, it's reduced the potential for loss of correction or implants they're shielding, the fatigue performance, reduce implant failure, the collapse of expandable cage and excessive wear debris evaluation or osteolysis. Sometimes migration testing of cages and total disc replacement, they reduce the potential for revision surgery and nerve compression. The need to perform fatigue testing to reflect intraoperative use of the implant should be considered whether it's an expandable cage or acutely bent rod.

I want to thank you for listening and I'm open for questions at this time. Thank you.

DR. HARNER: Thank you, Dr. Serhan.

Our next speaker will present on Industry Perspectives on Non-Clinical Performance Testing. Mr. Justin Eggleton is the vice president of spine regulatory affairs at MCRA. He is a former FDA orthopedic reviewer and has extensive knowledge of medical device regulations.

Mr. Eggleton.

MR. EGGLETON: Thank you, Dr. Harner. I'd like to start with a relevant quote from a

famous engineer, Scott Adams, who also happens to be the Dilbert guy. "Regular people say if it ain't broke, don't fix it. Engineers believe that if it ain't broke, it doesn't have enough features yet." I think that's an important quote to position this nonclinical testing discussion during a day where clinical studies and the data they generate is the topic.

Clinical studies provide a lot of meaningful data to define the safety, effectiveness, and the benefit-risk profile. However, the two do not operate independent of each other. Rather, regardless of the regulatory position, clinical data must exist to better inform the prescribed nonclinical test battery. I think a great example of this is total disc replacements.

If you look at the SSEDs from the first approvals in the mid-2000s and then the most recent, you can see the improved testing batteries, especially in the wear and the different edge cases, we've looked at the longer-term clinical data, especially explant studies. And this is the importance of simulating different scenarios from impingement and third bodies and their effect on the long-term wear profile.

Clinical studies themselves are insufficient to understand how devices perform and fail under various conditions. People are all different heights, weights, they move differently, they vary in muscle flexibility and joint mobility, and nonclinical testing allows us to model the worst-case conditions and to apply stress to the devices that may seem hyperphysiologic or just different.

As regulatory scientists, it's important for us to use these data to balance the premarket performance testing requirements and try to break everything down and categorize them into two distinct buckets. First we have nonclinical testing to demonstrate substantial equivalence in a 510(k). In other words, does the device type have an established clinical history whereby we know if it performs within specified boundaries, it can be considered substantially equivalent.

The second bucket is bookended by two parts. First, nonclinical testing to establish the reasonable assurance of safety to start the clinical study. Second, to contribute critical safety data to the safety discussion in the future marketing submission and fill holes where clinical data alone is insufficient. And another example here, picking on total disc replacements, fatigue testing, and wear testing. You know, the clinical studies can't be decades. At least, I hope not. So we can best model long-term safety through these more exhaustive conditions.

For the majority of spinal implants, I'd say nonclinical testing alone is sufficient for roughly 95% of devices. Within that 95%, I'd say 90% of those devices need bench testing only with the balance needing some type of cadaver or animal study.

Going back to all the devices, the other 5% are made up of Class III devices that require PMAs anchored by a clinical study, and then de novos and 510(k)s and more innovative devices that present different safety and effectiveness questions or device type that lacks an established clinical history.

You know, one issue faced quite often is side-by-side testing. In spine, FDA's done great work publishing acceptance criteria for cervical and lumbar interbody cages in the *Journal of Biomechanics*. There's also a 2020 guidance document for spinal putty systems that list out the acceptance criteria. And all these are a tremendous help to industry.

Industry, we hope this trend continues with the remainder of spine devices that require a 510(k) and that's because a lot of these devices require side-by-side testing, which is difficult for industry since these devices are prescription only and technically, it's illegal to purchase, you can't go to Amazon and get them, etc. We've heard pretty interesting ways around this. Distributors in other countries are willing to do a bunch paperwork or even swapping device samples at conferences. So certainly, it's quite an uphill battle to be able to procure a predicate device.

Despite these obstacles, the premise for the need can't be disputed, especially in those cases where non-standard testing is needed and without a comparator to a predicate device it's really difficult to position the results to demonstrate substantial equivalence. I don't like to propose a problem without a recommended solution; however, this is something I've struggled with for 20 years and I'll throw that out there again.

There is a role for clinical data in 510(k)s. For example, there's novel technologies or a risk that can't be evaluated on the bench, like I mentioned earlier. This can be inspired by a new failure mode that we see in a bench study or even lower test results that are outside the boundaries of the predicate devices. It can also be related to design input that's conceptually different and the modern testing techniques are insufficient to stress that particular endpoint. The scope of these clinical studies for de novos and 510(k)s are definitely device dependent and I'm a huge pre-submission advocate for these different scenarios.

Animal testing is generally useful for biocompatibility or material safety and sometimes also for functional purposes. I'd like to make the really bad joke that one of the downsides to working from home the past year and a half is that my dog won't talk to me after hearing how many animal studies we plan. But the importance and the contribution to the device safety profiles, it can't be ignored. The most common species we see are goats, sheep; several scoliosis models use mini-pigs, and rabbits are also used for material safety in more simple bone-healing response to synthetic bone -- those are generally -- fusion models since the rabbit disc space is too small for interbody fusion. For an interbody fusion study we generally see those done in goats or sheep.

Speaking to the functional studies first, they can be useful to evaluate a device's influence to the fusion process, particularly if there's a new material, surface treatment, or if it's accompanied by a drug or biologic. You know, functional studies do have their

limitations due to different biomechanics and signaling pathways, and also ethics play a role, especially when discussing subhuman primates. In discussing functional animal models, one must consider these differences and that's when we walk into a study design where the results may be biased due to a thin disc space or growth issues, particularly in scoliosis studies.

More importantly, animal testing has become the critical test to evaluate material safety. I looked through the ISO 10993 guidelines or to look at a more specific risk such as neurotoxicity. And I think neurotoxicity deserves a more open discussion. It's not explicitly covered by ISO 10993 or the FDA guidance, but it's a risk for devices in spine that are adjacent to sensitive important neural structures.

And lastly, speaking to the approximate cost of these different nonclinical studies, I want to share a few examples of common testing that we've seen. You know, a standard circle cage test battery to support a 510(k) can range from 30,000 to $40,000 and that's assuming dynamic testing at 10 Hz. Dynamic testing really drives the price, so if you need a test closer to 5 Hz or even 2 Hz, the price will go up.

Biocompatibility testing for a standard spine device that's bone contacting and categorized as permanent by ISO 10993, that will cost around $200,000 and it could take up to a year to complete. This timeline is really driven by implantation study time points and that can vary from device to device.

One important note to biocompatibility testing is for all the labs out there, their supply chain, cell lines, animals, regions, etc., were all hit really hard by COVID. That, plus an increased demand in response to newer FDA guidance has led to significant lead times, possibly several months or even more, so it's very important to engage these labs as early as possible.

A neurotoxicity study in rabbits with evaluations at 13 and 26 weeks, that generally

follows the Cunningham article, those can range from 250 to $300,000. The animal study itself, if fairly predictable. However, the cost of giant particles can certainly vary from material to material.

Cadaver testing also varies significantly compared to other tests since that scope is all over the place. A study with eight spines where devices are implanted at multiple levels, and there's some type of static in the TEG analysis, can range from 70 to $100,000. If it's a more simple surgical technique validation it could be half of that cost but again, it depends on how many spines and other factors, as well.

For MRI, the new guidance that FDA issued this year has been very helpful and there's an increased chance that technical rationales may be used to address many of the MRI-related safety concerns. Those rationales don't tend to cost more than a few thousand dollars, they don't typically require extensive testing. Simple devices like a simple plating system or a limited fusion device family are typically in the range of 15 to $25,000.

A more extensive product family like a thoracolumbar fusion system can be in the hundred to hundred fifty thousand dollar range, and spinal stimulators can easily acquire $250,000 for that body of testing.

Speaking to retrieval studies for explants, every analysis is highly specific to the device and the total cost can range from just a few thousand to over $20,000 depending on the level of analysis.

Tissue analysis through appropriately trained people who have the experience to interpret the results in a device-relevant manner can easily add $10,000 to the cost of analysis depending on the amount of tissue that's received. The critical step is to have a defined process for handling the devices in a -- scheduled fashion that complements the complaint handling process.

So that concludes my discussion of what I believe to be important discussion points

that contribute to the importance of medical device performance, whether nonclinical or clinical, as we seek to better serve the public. Thank you.

DR. HARNER: Thank you, Mr. Eggleton.

The next session will focus on the Overview of Spinal Device Clinical Review. It's my pleasure to introduce my colleague, Dr. Vincent Devlin, Chief Medical Officer in OHT 6. Dr. Devlin is a board-certified orthopedic surgeon with fellowship training in pediatric and adult spinal surgery. His professional interests related to orthopedic surgery include research, teaching, and participation in educational activities through professional orthopedic societies.

Dr. Devlin.

DR. DEVLIN: Hello, I'm Dr. Vincent Devlin, Chief Medical Officer in the Office of Health Technology 6, Office of Orthopedic Devices, within CDRH. Thank you for the opportunity to speak with you about clinical perspectives regarding the review of spinal device regulatory submissions.

This presentation is focused on three clinical areas related to spinal device regulatory submissions: first, the role of clinical evidence at various stages in the product life cycle for spinal devices; second, the different types of clinical data which device sponsors may choose to submit to the Agency; and third, it highlights the important aspects relevant to the review of spinal device regulatory submissions. In the sessions which follow, these key areas will be explored in more detail.

It is important to highlight that patients are at the heart of our work at FDA. As we review spinal device submissions, our focus is on how the devices can improve the health and quality of life of our patients. Our vision is for patients in the U.S. to have access to high-quality, safe and effective medical devices of public health importance first in the world.

To understand spinal device clinical trials, it is helpful to recognize that medical device trials are different from other types of clinical trials, especially drug trials. In general, device trials tend to enroll fewer participants than drug trials. Device and procedure modifications may occur during a trial. There is the need for flexibility in trial design and statistical analysis as large randomized controlled double-blinded studies may not be practical. For some devices, opportunities exist for leveraging alternative data sources such as existing registries or modeling techniques.

Clinical data may be needed at various stages along the total product life cycle for a spinal device. This includes during early device development and the subsequent clinical investigations performed to generate data to support marketing applications. The need for clinical data continues after devices reach the U.S. market so that device performance in the real world can be monitored and critical data can be communicated to patients, clinicians, and device developers. These data can also be integrated into development of next-generation devices.

Some of the scenarios across the total product life cycle where clinical data may be required are shown here. FDA's data requests follow a stepwise analytical process to ensure that the information requested reflects the least burdensome approach. When nonclinical test methods are not adequate or acceptable, clinical data may be requested by FDA.

With respect to device development, it is important to appreciate the distinction between the practice of medicine and clinical research. FDA does not regulate the practice of medicine. However, when clinicians use a significant-risk device which is not approved for a specific indication in the context of a clinical study protocol, this is different from practice of medicine and is referred to as investigational use or clinical research. In this type of situation, FDA approval of an investigational device exemption or IDE is needed.

There are three main types of IDE studies: early feasibility study, traditional feasibility study, and pivotal study. These various types of studies are intended to address different questions and collect different amounts of safety and effectiveness information. CDRH has several voluntary programs to encourage conduct of these studies in the U.S., as this helps fulfill our mission to promote innovation and new technologies for U.S. patients. These include the early feasibility study program, the Breakthrough Devices Program, and the Safer Technologies Program for medical devices and are described in detail on our website.

When device development progresses to the point where a marketing application is ready for submission to FDA, it is important to consider the regulatory pathways which are available. There are four regulatory pathways for marketing a device in the U.S.

The 510(k) pathway is most commonly used. To obtain 510(k) clearance, the submission must demonstrate that the new device is substantially equivalent to a legally marketed device called a predicate device. Although performance data are typically required for traditional 510(k)s, clinical performance data are requested only for a small number of submissions.

The premarket approval pathway is reserved for the highest-risk devices and requires the sponsor to provide valid scientific evidence to support a reasonable assurance of safety and effectiveness for their device. A PMA typically includes bench, animal, and clinical data to support claims made for the device.

The de novo pathway is used when the risk level of the device does not warrant the premarket approval pathway, but the sponsor and FDA are unable to identify a predicate device to support 510(k) substantial equivalence. Granting of a de novo classification request authorizes marketing in the U.S. and creates a new classification regulation.

The last pathway is the humanitarian device exemption pathway. This pathway is for medical devices intended to benefit patients in the diagnosis and/or treatment of a disease

or a condition that affects no more than 8,000 individuals per year in the United States. HDEs are exempt from the effectiveness approval requirement governing the premarket approval pathway but must show that the probable benefits of the device outweigh the risks, taking into account currently available devices or alternative forms of treatment.

Clinical data are also important after medical devices receive authorization for marketing in the U.S. CDRH utilizes multiple mechanisms to monitor device performance and device-related adverse events once devices reach the market. There are mandatory reporting requirements for manufacturers, importers, and device user facilities. Voluntary reports can be submitted by healthcare providers.

FDA recognizes the limitations inherent in these passive surveillance mechanisms and utilizes additional safety data sources. These include an enhanced surveillance network of approximately 300 hospitals, referred to as the Medical Product Safety Network or MedSun, as well as various types of postmarket mandated studies and compliance and enforcement activities such as FDA inspections of medical device manufacturers.

The most important recent development is the creation of a National Evaluation System for health Technology, referred to as NEST. Through NEST, access to large datasets from health systems collaborating together and from coordinated registry networks is opening up new possibilities for active surveillance of medical devices.

A wide range of clinical data sources may be used to support the safety and effectiveness of medical devices. The flexibility to use these various sources of clinical data is built into our regulations. The Agency relies on valid scientific evidence. What is acceptable under valid scientific evidence is shown here. I would like to highlight that it can include well-controlled investigations, partially controlled and uncontrolled studies, well-documented case histories, and significant human experience with a marketed device.

In addition to traditional pathways for clinical evidence generation, CDRH is working

to increase the use of observational data. Many points of the healthcare system interface with medical devices as they are used by patients and clinicians and can generate additional data. FDA is increasing the use of data generated from registries, electronic health records, and claims data and leveraging this data to monitor postmarket safety and adverse events that inform regulatory decisions.

A good place to start to answer questions about real-world evidence for medical devices is to look at the 2017 CDRH guidance document on real-world evidence. This document highlights the importance of data quality in terms of relevance and reliability and highlights the importance of patient protections. Examples of how real-world evidence has been used in regulatory decisions for spinal devices are available in this publication on the CDRH website.

Here are the focus areas relevant to the clinical review of spinal device regulatory submissions that will be discussed today. It is important to adequately describe the target study population. Attention should be directed to ensure that inclusion and exclusion criteria are consistent with the desired target population. There should be a balance between strict definition of the ideal study population versus inclusion of the types of patients that will benefit from the spinal device under real-world conditions. If a control group is included in the study, there should be comparability between the test group and the control group to minimize bias.

Individuals are identified as enrolled in a clinical trial following completion of the informed consent process, and attempts should be made to collect data on reasons for subsequent non-participation and efforts taken to minimize loss of follow-up.

Common clinical assessments in spinal device studies include evaluation of pain, function, neurologic status, subsequent surgical interventions, adverse events, and imaging outcomes. Success is most often evaluated through at least the 24-month time point and a

minimum of 85% follow-up is recommended to maintain the power of the study. It is critical to conduct the right assessment at the appropriate time as one cannot go back in time to collect all relevant data.

Clinical outcome assessments can be reported by the patient, a healthcare provider, or a nonclinical observer or through performance of an activity or a task. The outcome assessments that will be discussed over the course of today in relation to spinal devices include patient-reported outcomes, clinician-reported outcomes, performance outcomes, and imaging assessments. The preferred imaging modalities and assessments depend on the patient population, device type, and study goals.

In spinal device regulatory studies, it is recommended that sponsors submit an imaging protocol which outlines standardized imaging assessments and definitions, as well as measures to minimize bias. Additional recommendations include the use of two independent imaging readers and an adjudicator for any disagreements and for large trials, use of an independent core imaging lab.

For spinal device studies, the safety parameters which should be included are adverse events, subsequent surgical interventions, neurological adverse events, and deaths. Regarding adverse events, a pre-specified plan is necessary for collection and categorization of adverse events with respect to seriousness, severity, relatedness to the device or procedure and whether AEs are anticipated. Statistical analyses and narrative descriptions of major safety outcomes are requested.

Subsequent surgical interventions and neurological adverse events should be recorded and categorized, and a risk-based monitoring plan should be created and may include a clinical events committee and a data monitoring committee.

When evaluating and reporting spinal device effectiveness, parameters which should be included are pain, function, health-related quality of life, and the other parameters listed

here.

Study endpoints may be considered as primary, secondary, or tertiary. Primary endpoints are designed to evaluate the safety and effectiveness of the device, and for spinal device studies generally include assessment of pain, function, subsequent surgical interventions, and major complications. Secondary endpoints evaluate additional meaningful claims and provide further insight into the device effects and mechanisms of action. Tertiary endpoints investigate new hypotheses or clinically important events that occur infrequently.

Primary endpoints should evaluate the safety and effectiveness of the device in the population in which the device is indicated. Spinal device studies generally evaluate multiple endpoints. When there is more than one primary endpoint a study is considered successful if a subset of the multiple endpoints is met.

There are different types of multiple endpoints. In the case of co-primary endpoints, multiple safety and effectiveness endpoints are individually evaluated and the study is considered successful if all endpoints are met. In the case of composite endpoints, several clinically relevant endpoints are combined into a single endpoint. Composite endpoints are commonly used in spinal device clinical trials.

Here is an example of a composite primary endpoint which illustrates the combination of various clinical outcomes into a single variable or composite endpoint. It is important to define the relationship between individual patient success and study success, and predefine the success criteria for each. Time course distributions for the individual endpoint parameters should be reported at predetermined study time points.

Regarding statistics for spinal device regulatory studies, the statistical analysis method should be described in the clinical study protocol and the statistical analysis plan finalized before data become available for analysis. A sample size justification should be

provided. Recommended data presentations include time course distributions for important evaluation parameters, as well as narrative descriptions for major safety outcomes.

CDRH places emphasis on benefit-risk assessment as the central component of the medical device evaluation process. Listed here are some factors which are considered in making a benefit-risk determination for a specific device.

CDRH considers additional factors in the determination of benefit-risk including the degree of uncertainty of the benefits and risks of a device and whether the device is a novel technology addressing an unmet need.

By recognizing a measure of uncertainty and by putting our patients first, we have been able to construct a benefit-risk framework for medical device decision making and better understand what patients want, the risks they are willing to tolerate, and what they value as benefits.

Here are some key takeaways I would like to end on. Robust clinical evidence is integral to the marketing authorization of higher-risk and innovative medical devices, but the most appropriate and least burdensome pathway to market varies based on all of the factors I have discussed. For some technologies intended to address important unmet needs, it may be appropriate to accept a greater degree of uncertainty about the safety or effectiveness of the device and consider patient perspectives regarding tolerance for risk as well as their perspectives on benefit. Also, it is critical to ensure that clinical trials are designed to assess what matters most to patients and to facilitate diverse patient enrollment in clinical studies.

Thank you for the opportunity to speak with you today.

DR. HARNER: Thank you, Dr. Devlin.

Next we are going to shift gears a bit and hear a presentation on the Strategies for

Enhancement of Diversity in Spinal Device Clinical Trials. Dr. Sydney Gibson, a bioengineer/reviewer in the Division of Spinal Devices in OHT 6, will be presenting on behalf of Dr. --

UNIDENTIFIED SPEAKER: Dr. Harner, you're muted.

DR. HARNER: Oh, sorry. We're going to hear from Dr. Sydney Gibson, a bioengineering reviewer in the Division of Spinal Devices in OHT 6. She will be presenting on behalf of Dr. Elizabeth Panox, who is an acting deputy director in OHT 6.

Dr. Gibson.

DR. GIBSON: Good morning, my name is Sydney Gibson and I'm a lead reviewer in the spine division and I will be presenting on behalf of my colleague, Dr. Elizabeth Panox, today and I'm here to discuss strategies for enhancing diversity in clinical trials. To set the foundation for this talk, let's first review how we define a disparity.

Disparity is often used interchangeably with inequality and inequity, but there's no universally accepted definition of disparity and in fact, 11 different definitions exist. However, the choice of definition may include the choice of comparator, a policy implication, and/or a sense of unfairness or injustice.

There are three broad approaches to defining a disparity: using a comparison with a non-minority population, using a comparison with the general population, or observing differences among different population segments, though with this approach the comparator group may be unclear.

The common definitions of disparity that are used come from the Institute of Medicine, the National Institutes of Health, and Health and Human Services Healthy People Initiative.

In 2002, the Institute of Medicine defined disparities as racial or ethnic differences in the quality of healthcare that are not due to access-related factors or clinical needs,

preferences, and appropriateness of intervention. Later, in 2010, the Healthy People Initiative defined disparity as a particular type of health difference that is closely linked with social, economic, or environmental disadvantage. In 2015, the National Institutes of Health defined it as a health difference that adversely affects a disadvantaged population.

There are many sources of disparity and different factors can contribute to an individual's ability to achieve good health. In addition to race and ethnicity, disparity can arise from differences in sex, sexual identity, age, disability, socioeconomic status, and geographic location.

Disparities can have adverse effects on the health of racial minorities and people of color. People of color experience higher incidences and higher severity of many conditions such as diabetes, end stage renal disease, cancer and more recently, COVID-19. They can have greater severity in presentation in many conditions. They are also less likely to undergo routine screening, have poorer treatment outcomes, and they are less likely to be offered advanced treatment for their conditions.

In orthopedics, many disparities can be observed for racial minorities and people of color including higher incidences of osteoarthritis and differences in rates of hip and knee arthroplasty. Additionally, these groups are more likely to have surgery delayed and more likely to have complications following surgery.

Orthopedic surgeons do acknowledge the existence of disparities. Most believe that the cause is a lack of insurance and about a quarter believe that research into disparities would be very useful.

And going even further into orthopedics, racial disparities are also present in spine care. We see this in African American teenagers with adolescent idiopathic scoliosis who tend to present with curves of greater magnitude. After lumbar spine surgery in adults, race is an independent risk factor for complications, and African Americans with spinal cord

injury are less likely to receive surgery and have a higher rate of morbidity and mortality.

Ethnic and racial disparities are complex and multifactorial in etiology and so to close a disparity gap, a multipronged approach is needed and one strategy is to increase minority participation in clinical trials. Racial and ethnic minorities have been historically underrepresented in clinical trials and we need representation in trials to study the effects of medical products in the people who will ultimately use them. Devices may perform differently in persons of different ages, races, and ethnicities. This also means that participants can react differently to certain medical products based on their background. We can better understand health disparities when we look at diseases that occur more frequently or appear differently in diverse populations.

There are many barriers to minority participation in clinical trials including distrust of the medical system due to historical abuses, a lack of awareness of what a clinical trial is and what it really means to participate, and as well as inadequate recruitment and retention efforts of minorities into trials. Barriers can also arise from a lack of cultural understanding such as misunderstanding minorities' beliefs and values that contribute to their decision making process or a lack of culturally and linguistically appropriate communication or the damaging perception that racial and ethnic minorities don't want to participate in clinical trials. Trial design such as enrollment criteria and time constraints can also impede enrollment of diverse participants.

The Healthy People Initiative is led by the Department of Health and Human Services and establishes 10-year national health objectives. It's an expanding and an evolving program that collaborates with many federal agencies, specialty societies, and academia. The healthy initiative defines a disparity as a particular type of health difference that is closely linked with social, economic, or environmental disadvantage. The overarching goals of the program have evolved over the past 30 years to accommodate the expanding and

changing definition of public health. And over three decades ago, in 1990, the program set an initial goal to decrease mortality in children and adults, and to increase independence among older adults. And the goals have evolved since then. Most recently, in 2020, the Healthy People Initiative's overarching goals are to attain high-quality longer lives free of preventable disease, achieve health equity, to create social and physical environments that promote health, and to promote quality of life and healthy development.

However, the most recent status report from the Healthy People Initiative shows us that there's still work to be done in gaining health equity across racial and ethnic groups. When observing which groups experience an improved quality of life, disparities remain and there's still work to be done in order to address these gaps.

The FDA has taken actions to address racial and ethnic health disparities through legislation and guidance for industry and FDA staff. The FDA's Safety and Innovation Act, Section 907 action plan recommends that medical product applications submitted for marketing approval improved the demographic subgroup data's quality, completeness, and availability.

The plan focuses on three priorities: improving the completeness and quality of demographic subgroup data, encouraging greater participation by employing target strategies and identifying barriers to subgroup enrollment in clinical trials, and transparency through improving the public availability of demographic subgroup data. These priorities aim to encourage the inclusion and greater representation of a diverse patient population in biomedical research leading to development of medical products.

The 2016 FDA guidance document "Collection of Race and Ethnicity Data in Clinical Trials" provides recommendations regarding how race and ethnicity data is collected. One recommendation is for trial participants to self-report race and ethnicity information and those individuals be permitted to designate a multiracial identity. This can allow for better

accuracy in identifying patient race and ethnicity, but it can be challenging to obtain discrete pre-identified categories.

Another recommendation is to use the two-question format for requesting race and ethnicity information with the ethnicity question preceding the question about race. Using the two-question format can allow for better consistency in data, but there isn't always a consensus on the definitions of race and ethnicity within a given community. For example, a participant would be asked whether they identify as Hispanic or Latino before choosing the racial designations that best describe them. The ethnicity category of Hispanic and Latino as defined in the guidance is shown here. However, not every participant within this community would identify with this definition of their ethnicity.

The FDA recommends the following five minimum race categories be offered: American Indian or Alaskan native; Asian; black or African American; native Hawaiian or other Pacific Islander; and white. And their definitions are detailed in the 2016 guidance.

In certain situations more detailed race and ethnicity information may be desired and can enhance understanding of the trial participants. For example, for clinical trials conducted outside the United States, the recommended categories for race and ethnicity may not adequately describe groups in foreign countries. Furthermore, some categories can reflect origins from large or widespread areas. So for example, Asian can spread from India to Japan. In situations where appropriate, the FDA recommends using more detailed categories by geographic region to provide sponsors the flexibility to adequately characterize race and ethnicity.

However, limitations still exist when defining racial and ethnic categories. Some confusion remains about the distinction between race and ethnicity and when considering ethnicities, Hispanic is often reported but there are numerous distinct ethnic groups that aren't always included. There are currently no provisions for reporting people of mixed

race and some terminology used can be offensive to subjects. The guidance provides recommendations for collecting and reporting race and ethnicity data in clinical trials, but gaps remain. The guidance doesn't provide recommendations for data presentation for marketing applications submitted to CDRH, it does not address what level of participation is recommended in clinical trials, and there are no recommendations for analyzing race and ethnicity data.

The 2017 guidance document "Evaluation and Reporting of Age, Race, and Ethnicity Specific Data in Medical Device Clinical Studies" recommends that sponsors consider any known or potential age, race, and ethnicity specific differences when designing a study protocol. This can include differences in disease course, outcomes, or benefit-risk profiles among groups. It's recommended that this information is included in marketing applications, interim reports, final reports, and eventual labeling.

To enhance the recruitment of relevant subgroups, various approaches can be considered including having a variety of investigational sites where recruitment can be more easily facilitated or revising the enrollment criteria, when appropriate. They can also provide flexibility in follow-up, for example, allowing patients opportunities that match their schedules to include evenings and weekends, leveraging local healthcare practitioners and minority health professional organizations to refer patients, or offering incentives, for example, providing compensation for transportation costs.

In clinical trial reporting and analysis, the race and ethnicity composition of the study population should be reported. Subgroup analysis based on race and ethnicity should also be included. If the proportions of participants enrolled are substantially different than the prevalence of disease by age, race, or ethnicity, the generalizability of study findings should also be discussed during reporting. Differences in study outcomes among groups should also be identified. All of this information should be included in publicly facing

documents.

There are best practices that clinical trial sponsors and investigators should consider which can help avoid or minimize loss to follow-up of subjects and improve diverse participation throughout the duration of the study. Trial sponsors should consider producing study protocols that include provision for capture in analysis of ethnicity and racial data; develop follow-up plans that detail follow-up goals, frequency of scheduled visits, proxy contact information, and actions to be taken when a subject misses a visit; demonstrate continued interest in the study subjects; monitor follow-up rates closely; and report subject accounting data in study reports.

Clinical trial investigators can also engage in best practices to help achieve a diverse clinical trial enrollment and retention. They should consider participating in cultural competency training prior to study recruitment. They can counsel subjects on the importance of follow-up during the informed consent period and thereafter. They can remind subjects of follow-up visits and can attempt to locate subjects who miss their follow-up visits and encourage them to return. They can obtain proxy contact information when unable to contact a study subject. They can ask subjects who withdraw to provide a reason why and to assess their experience. And they can provide close follow up with subjects, for example, having a telephone call follow-up after surgery, particularly for implantable devices.

We've talked a lot about considerations for sponsors and investigators, but here in the Office of Orthopedic Devices, we are also spearheading efforts to enhance clinical trial diversity. We've completed an internal SOP best practices document, written for feedback to sponsors and for review of best practices, we ensure that IDE protocols include a strategy for recruitment and retention of minority groups, and we have updated device labeling where possible to reflect performance in minority groups. We are in progress for adopting

an internal plan for reversing disparities and we're currently collaborating with specialty societies and other external stakeholders in educating patients and providers. Lastly, there's ongoing research to characterize the extent of disparities in FDA trials.

In the future, our office intends to develop guidance to adopt a definition of disparity, continue research into trends of disparities, and continue outreach to external stakeholders and patients.

With that, we'd like to acknowledge the Office of Orthopedic's leadership, Captain Peat, Dr. Devlin, and Dr. Jean, as well as the Office of Minority Health. And we'd also like to thank our team members and summer students who have contributed to OHT 6's ongoing efforts, including Dr. Kavlock, Dr. Jiang, Cecilia Groves, Marissa Marine, and Joseph Peluso.

DR. HARNER: Thank you, Dr. Gibson.

The next presentation, Analysis of Diversity in Spinal Device Clinical Trials, will be delivered by Dr. Hongying Jiang, a clinical epidemiologist and safety signal manager in OHT 6. She has been with the FDA since 2011 and has led epidemiology teams on post-approval and postmarket surveillance studies.

Dr. Jiang.

DR. JIANG: Good morning, everyone. My name is Hongying "Helen" Jiang, a health scientist and the safety signal manager in OHT 6. Today I would like to present our preliminary results about diversity research in spinal device clinical trials on behalf of our OHT 6 team, Dr. Kate Kavlock, Dr. Panox, and Captain Peat.

First, we've collected all baseline demographics from clinical trials in all OHT 6-approved PMA, HDE, and de novo submissions. There we've conducted descriptive analysis including various counts, calculating minimum, medium, maximum, and percentages of the study subjects in any study arm for each subgroup. We've also plotted demographic subgroup data over time and compared to the racial and sexual demographics of the

general U.S. population as reported by the U.S. Census Bureau from 1990 to 2020. Next, I would like to show you some of our preliminary results. Next, please.

Figure 1 is a flowchart that shows the steps we've taken to reach the final dataset. There are a total of 94 PMAs, HDEs, and de novos. Each submission may have more than one cohort or trial or study, for example, an IDE cohort and a continued-access cohort. Therefore, we've collected a total of 254 records of cohorts. For today's presentation, we are focusing on analyzing the investigational or device arm that has more than 30 patients, which means 128 and 17 studies were excluded. This left us with a total of 109 studies or cohorts. A quick breakdown among three of our divisions, divisions A and C had 37 and 38 studies, respectively, and the spinal division B had a total of 34 studies which are the subject of the following analysis. Next one, please.

In the final 109 device groups, we've included a total of 22,000 patients and the median enrollment was 151 patients per study group. Within the total studies, mostly 103 were PMAs, five HDEs, and only one de novo. Within the 34 spinal device cohorts, we have included 5500 patients and the median enrollment was 164 per study, a little bit more than other ortho devices and again, mainly PMAs, a total of 32 PMAs and two HDEs and no de novos.

Let's look at some figures. Figure 2 is the average female enrollment percentages in spinal devices grouped by five device types. X-axis is the percentages, Y is the five groups including parentheses with the number of studies reported sex enrollment over the total studies per study type. For example, in the last bar here there were two studies for idiopathic scoliosis (IS) correction devices, both of them, two, over two, reported female percentages. From this figure we can see that the top two types had average enrollment around 40%. The middle two were 50%, but the last two IS correction devices had the highest, around 88%, which we believe reflects the fact that female patients are the

predominant sex in the real-world situation. The overall percentage is around 51 to 52%, which is pretty even between the two sexes and range from 38 to 89%. This number appears to be higher and broader than other device areas.

Of note, on a total, disc replacement devices had 21 studies. All the rest of device types had very limited 1, 2, 4 studies, therefore the following ratio distribution analyses were based on those 21 total disc replacement devices and due to high missing values on certain racial groups, we're only presenting the white, black, and Asian American percentages in the next two slides.

First, let's look at reported white Caucasian percentages over time. Here, the X-axis is the years since 2000 and the Y is the enrolled percentages. Each dot represents a total disc replacement device group, a study, and the red dots represent three census data reported in year 2000, 2010, and 2020. And the dotted line is the linear trend line for the census data. Apparently, the white was enrolled between 80 to 100%, which were overrepresented when compared with the percentages by the National Census data. The minimum, maximum, and median sample sizes for all the studies in here are provided for your reference on the right. Next slide, please.

Same figures were made for the black African Americans and Asian Americans. Please note that the Y-axes are much smaller than the previous figure, 3A. Apparently, they were both underrepresented. Specifically, blacks were enrolled between 1 to 8% and Asians were enrolled less than 5%. They are much lower than the corresponding census data. The minimum, maximum, and median sample sizes were also provided to the bottom of each figure for you to see the specific patient numbers.

In summary, from the above preliminary analysis, we can see that there are significant sexual and racial disparities in spinal device trials. For example, female enrollment ranged from 40 to 90% and the white was overrepresented, while the black and

Asians were underrepresented. Some disparities appeared to be related to the underlying disease distributions. Additional investigation is warranted to identify other disparities and various challenges in enrolling underrepresented patient populations. And finally, additional efforts should be encouraged to enroll underrepresented patient populations in spinal device clinical trials. Next, please.

Thank you for your attention. This ends my presentation.

DR. HARNER: Thank you, Dr. Jiang.

We will now start our first break, it's a 10-minute break. It is 10:14, but we will resume at 10:25.

To begin the session after the break will be our first clinical discussion. Please send your comments and questions to OHT6-Feedback@fda.hhs.gov. Thank you and have a great break.

(Off the record at 10:14 a.m.)

(On the record at 10:25 a.m.)

DR. HARNER: Welcome back. We will now begin our first clinical discussion related to Defining a Target Population, Enrollment Criteria, and Strategies for Inclusion of Underrepresented Groups. Please join me in welcoming the following five nationally recognized spine surgeons.

Dr. Paul Anderson is a Professor of Orthopedics at the University of Wisconsin School of Medicine and Public Health. Dr. Anderson is a nationally recognized expert in spinal trauma and complex cervical spine disorders, with research specialties that include spinal fixation implants and the development of a spinal cervical disc replacement.

Dr. Darrel Brodke is a Jack and Hazel Robertson Presidential Endowed Professor and focused on spinal biomechanics, spinal deformity surgery, treatments, and patient-reported outcomes.

Dr. Raymond Golish is the chief medical officer at HCA Healthcare at JFK Hospital in Palm Springs, Florida. As a surgeon, scientist, and healthcare executive, he is consulted widely at the forefront of industry, government, and professional societies. He has served in leadership roles through the academy, NASS, HIMSS, and ASTM.

Dr. Khaled Kebaish is the division chief of spine surgery and Professor of Orthopaedic Surgery and Neurosurgery at the Johns Hopkins University School of Medicine. Dr. Kebaish has research interests in clinical and functional outcomes of spine surgery, with a special focus on adult deformity surgery, as well as an interest in biomechanical testing and evaluation of different fixation devices and techniques in the adult spine.

Dr. Noelle Larson is Associate Professor of Orthopedics, Director of Research for the Division of Pediatric Orthopedics and Scoliosis, and a pediatric orthopedic surgeon at the Mayo Clinic. Dr. Larson's clinical interests include spine deformity, early onset scoliosis, and neuromuscular disease. She has held leadership positions in the Pediatric Orthopaedic Society of North America.

Next, I would like to introduce two of our medical officers from the Division of Spinal Devices within OHT 6.

Dr. Dirk Alander recently joined the FDA after serving as a spine surgeon and chief of quality for the Musculoskeletal Institute at Geisinger Health System in Danville, Pennsylvania.

Dr. Caroline Moazzam joined the Agency 4 years ago after working in clinical practice in Northern Virginia.

I'm going to ask Dr. Alander to please introduce the first topic.

DR. ALANDER: Good morning. The first topic for clinical discussion entails health disparities across adult orthopedic spinal conditions and the adult spinal device clinical studies. The questions for discussion are:

How can we increase representation by underrepresented groups?

What are the biggest challenges to inclusion of the underrepresented groups, and how do we most effectively address these issues?

And lastly, how do we account for the fact that different diseases vary across populations, for example, prevalence of osteoporosis that is influenced by ethnicity and race?

DR. DEVLIN: If Dr. Kebaish could please go ahead and introduce the session.

DR. KEBAISH: Okay. Well, thank you for having me on. Clearly, this is a very important topic that affects how we take care of our patients because what we do at the FDA in researching devices and subsequently identifying the best way to introduce those devices and use it in patients, and one of the concerns clearly is how to properly investigate those devices and be representative of the entire population and not a select group or -- and especially where we have the diversity of disease, we don't necessarily have the same representation when those devices are being tested and investigated.

So I think this is something that we need to do better on. We've had much research, including some recent publications from our group looking at the impact of the lack of disparity and how care delivery -- which starts by doing our job early on and I think the FDA has been trying and improving on this particular concern.

I think we would -- so one of the questions is how do we increase representation by underrepresented groups? Because we all know there are some concerns in patient participation in those, whether it's because of mistrust or because of our inability to reach those groups and incentivize them to be part of our FDA trials that are ongoing. Maybe I'll start by putting that question out to the panel and let me start by asking Dr. Brodke.

DR. BRODKE: Thank you, Dr. Kebaish. It's particularly a challenge in my state where we are known not to be as ethnically diverse as many, although we do have a large

population of Pacific Islanders and Hispanic ethnicity.

The challenge is twofold. One is a challenge of having the right mix when it comes to enrolling patients into studies, and the second is the issues brought up earlier and well discussed in the presentation this morning around trust for certain ethnicities to the medical research process, and I think we have to address both of those head on to be able to really meet this challenge.

And so thirdly, I guess maybe just from a statistical representation standpoint, we should be discussing how to manage enrollment so that we make sure that we get the broad mix of race and ethnic diversity that we're looking. Not always easy when we have subsets that are small enough to challenge our statistical power to understand what is happening and what we're testing.

All of those, I guess, here -- all of the three things I just discussed really underscore the challenges we face and not really the answers to those challenges. I think it takes a specific focus on a part of the investigators. It may require some change in study design, specifically as it relates to race and ethnicity, and I think we probably should be talking about those things in specific. Maybe I'll stop here and let others comment. Thanks.

DR. KEBAISH: Thank you, Dr. Brodke.

Dr. Anderson.

DR. ANDERSON: Yes, thank you. I've been thinking about this since Dr. Devlin invited me to participate and having participated from Wisconsin, which is also, as Darrel just said, not a racially diverse community, is how do we include more diversity? Well, I think you probably need to broaden the inclusion/exclusion criteria and one of the problems of FDA trials, as they try to generalize to the population, is that the trials are very narrow and actually 95% of our patients who come in don't meet the criteria for the trial anymore when we start using the device. So having broader inclusion and exclusion criteria

may help. As an example, osteoporotic trials are always started in women and exclude men, so drugs get approved for women but not for men. And in fact, 25% of the people with osteoporosis happen to be men, so it's not just a woman's disease.

I think you need to particularly target diversity populations when there is a known biologic reason for that and what I mean is that there are genetic factors that predispose to either the disease itself or complications related to the device and if those are known, then you really need to carefully power the study to include that ethnic diversity. So those would be a couple of things.

The other one is to look at the locations where the studies are done. Typically, most of the -- a lot of these FDA trials are done at academic medical centers or in large private practices that are very busy, usually in suburban areas, and so they're not getting a lot of the poorer people from the urban populations. And so it would be incumbent to make sure that there is diversity in the patient populations of the investigators, which would hopefully drive better diversity in the clinical trials. Those are my thoughts.

DR. KEBAISH: Thank you, Dr. Anderson.

Maybe I'll move to asking Dr. Larson. I mean, even here I would say, here in Baltimore we have a very diverse population but still, I can argue that the access is still not ideal and if patients don't have access to the service or the physicians that are participating in these trials because, whether it's insurance or the particular physician is not seeing that patient with special insurance or maybe underrepresented minority who have -- don't have the full access. So how do we overcome maybe the economic factors that are impacting or negatively impacting -- or proper presentation of under-minority groups?

DR. LARSON: Just a pleasure to be here today and I was just going to add to Dr. Anderson's comments. Many of the barriers are economic and I think additional support for hospitals which are under-resourced, who may not have the ability to do a trial

or have never done a trial before, so additional support so that centers, again, that aren't major academic centers can successfully be a trial site.

I'd say also, additional resources for the patients who enroll. Many of the studies I've been a part of will provide the patient with $25 or $50 for coming to a study visit. But really, by the time you think about transportation, parking, missed day of work -- for pediatrics, missed day of school -- there is a significant volunteer effort for patients to be part of an FDA trial and trying to compensate that for patients who are not well resourced and may not be able to afford time off work.

Also, I think creative follow-up strategies such as virtual visits or mailed-in X-rays that don't require such intensive resources on the part of the family to participate, I do think that's a key part to recruiting and actually keeping a diverse group of patients populating the study and following through the study to 2-year follow-up or whatever the target time frame is.

DR. KEBAISH: These are excellent comments and clearly highlights the need to figure out the economics of underrepresentation in some of the trials.

Dr. Golish, any comments on your experience and do you have some comments on what you think would allow us to improve the process?

DR. GOLISH: Yeah, I'd like to start with the economic disparities and challenges and talk on behalf of sponsors I've been associated with. You know, sponsors, in running a big study, have a large economic burden but also large resources to match that challenge to get the highest quality of evidence, especially for pivotal IDE trials for Class III devices. In that context there's money on the table, but the notion of subsidizing individual sites and individual investigators who have access to diverse populations is appealing. I've heard it be appealing to sponsors. However, they have a legitimate compliance risk in even contemplating that reality. So I think since FDA has done the good work of inviting fresh

thinking around this topic, thinking about how compliance could look with sites that have access to underprivileged populations and will do the hard work in order to recruit those individuals, that would be something worth meditating on the mountaintop about.

And that brings me back to the first point which both Paul Anderson and Darrel Brodke mentioned, which is the level of trust and that is very real. So I've had the advantage of being a healthcare executive and administrator throughout a period of pandemic and got to see thousands of underrepresented people's medical decision making up close, and the fact of the matter is they have sometimes told me -- it's not my opinion, but just me reflectively listening to things, people have had the courage to tell me that they do have a significant mistrust in medical, scientific, and governmental authorities and establishments and I think the only way to get through that is grassroots rather than grass top, is to have individual investigators who have access to these populations, be they underrepresented investigators themselves or otherwise, to put in the hard work, one on one, patient by patient, in order to build a rapport and break through that trust of a human-to-human connection rather than some kind of statistical or policy aggregate.

I think perhaps those two things, the grass tops with some kind of compliant mechanism to incent underrepresented populations and investigators, but then simultaneously a grassroots effort to break through that barrier of trust would be the two leaves of bookends.

DR. KEBAISH: Dr. Golish, I mean, the trust issue, I think that's really important and I think, in Baltimore here, we've had this concern and whenever I mention a study to a patient and specifically underrepresented, they always think of the history that we all know from taking advantage of some population. So I think we have a lot of work to do in that area to change that and to bring back trust and that this is not an experiment to try to take advantage of people and experimental as it has happened, you know, in the past. Thank

you.

Do we still have time for additional comment? Any other comment?

DR. HARNER: Yes, we're going to go to 10:45, Khaled.

DR. KEBAISH: Okay, great. Dr. Devlin, any thoughts on this topic?

DR. DEVLIN: I would agree with all of the points made and just emphasize that education is very important and that will help break down the barriers, I think education not only of patients, but also the investigators, and just through having meetings like we're having today helps focus attention on this area and helps us to keep this important topic in our mind, in addition to all the other challenging topics we need to be aware of in a busy clinical practice and busy clinical study.

DR. KEBAISH: On the education topic, I mean, it's clearly easy to educate clinicians and those who are involved in these trials, but how do we reach the target population with education? Is that through media, through advertising? How do we -- because that takes a lot more work and it's not as easy as we think, especially knowing that we need to have a much wider audience and we need to reach them where it's most appropriate and do we do it through venues that we know they are most likely to trust and attend or how do we do better in educating the public?

DR. ANDERSON: I have a suggestion, is that since this is a generalizable --

DR. KEBAISH: Okay.

DR. ANDERSON: -- problem for the FDA in medical trials of all types, perhaps they should create educational material for potential subjects enrolled in this that addresses the issue that we, as investigators, can refer patients to, who have some questions about trust.

DR. DEVLIN: I think that's a great suggestion, Paul, and certainly tailoring the education materials in terms of language and focusing on what would help the specific patient group that's being outreached to, certainly can contribute.

DR. LARSON: I could share from my experience. We recently completed a randomized controlled trial for children with scoliosis for more screws versus fewer screws, and one of our pilot studies was to try the educational module for about 200 families and we asked the families, after going through our educational module, "What would make you be willing to enroll in the study?" And we found that with our initial educational module, only 70% of patients understood what randomization was and the families that didn't understand what randomization was were less likely to want to enroll in the study.

So it really seems that if we're doing a formal consented prospective study, having that education piece either through the study coordinator or through the surgeon, that's very regimented and has been reviewed by a patient education specialist to make sure we're presenting it at the proper level with the proper perspective.

I'd say another kind of left turn you could take is the whole concept of real-world data and collecting data as part of clinical care and that could help for a more diverse enrollment strategy because everybody understands clinical care and particularly in the role for postmarket surveillance, collecting information about something that's already happened is a lot easier for families to accept. Being in a registry is easier than being randomized for a study. So I think that move of the FDA towards real-world data is also powerful for recruiting under-represented groups.

DR. HARNER: Thank you.

DR. KEBAISH: Thank you.

DR. HARNER: Excellent discussion. We need to move on to the second topic, which is also related. So Dr. Alander's going to pose three more -- two or three more questions for the panel related to this topic.

DR. ALANDER: Very good. Our second topic for the panel discussion involves health disparities across the pediatric orthopedic spinal conditions and pediatric spinal device

clinical studies. The questions, much like the first:

How do we increase representation of underrepresented pediatric groups?

What are the challenges to inclusion of underrepresented pediatric groups, and how do we most effectively address these issues?

And lastly, how do we account for the fact that different diseases vary across populations and are influenced by ethnicity and race?

DR. LARSON: I think I'm going to lead off the topic as a pediatric orthopedic spine surgeon. And this has been a great discussion so far today and really a true honor to be in this esteemed group. And as we've heard, it's not just racial or ethnic differences, but also socioeconomic differences and I'll add sexual and gender minorities, as well, to the list.

The difference with pediatrics is we're looking at a proxy decision-maker situation, so it's not the individual deciding to be in the trial, it's a parent or guardian or legal representative and that's adds, I'd say, a whole other level of kind of ethical considerations.

I think the other thing with pediatric spine conditions is many of these conditions occur in conjunction with other very severe problems like cerebral palsy or spinal muscular atrophy or a very heterogeneous group represented by small patient populations distributed over large geographic areas.

Particularly for the neuromuscular patients, I mean, many of these families, the parent has quit working to care for the child, so they're on public assistance and have limited funds to pay for travel and adding a clinical trial or a study on top of their already fairly burdensome schedule is a challenge.

For pediatrics in general, we looked at the brace trial data, so this was a prospective, randomized controlled trial of children being randomized to a brace versus no brace, based in Iowa, and we showed that children who were from a low socioeconomic zip code wore their brace on average three hours less than children from a better zip code. So for sure,

there are disparities that affect the treatment of spine deformities and we really need to include all this as we consider designing trials and consider enrollment in trials.

So I think I've kind of hit on some of my already favorite points here, as far as like real-world data is going to help rare diseases, collecting data through registries will be less burdensome. I think the role of virtual visits and mailed-in radiographs in my practice has played a huge role. I think, previously, I was bringing patients back for clinical follow-up at 3 months, 6 months, 1 year, 2 years, 5 years, and now at least half of those visits are being done virtually. The radiographic quality is not as good. I'm still getting outcomes questionnaires. The exam is not as good. It's clearly not good enough for a high-level prospective trial but again, for the sake of recruiting more patients, patients who represent the general community and also reducing the burden on families, maybe we need to consider more of these virtual visits for follow-up.

And again, like we've heard earlier, it's the Hawthorne effect, right? If we shine a light on this problem, we're all going to be thinking about it, we're all going to consider it as we design our trials. More and more journals now are requiring analysis of gender differences, racial differences as part of the final analysis and more and more studies and funding organizations are requiring it. So I think if we hit the problem at every level in the spectrum, from study design to publication, that's also going to help us shine a light on it and show people that we value having a diverse group.

But maybe some of the other investigators that have actually run studies can speak to anything about pediatric patients or rare diseases or conditions that aren't as common as some of the more adult degenerative conditions.

Dr. Brodke?

DR. ANDERSON: I'll just make a comment about outside of adolescent idiopathic scoliosis, I think, in pediatric spine deformity, the number of patients is going to be so small

that it's not feasible to do randomized controlled trials and consequently, high-quality observational study designs are probably what's needed with reasonable longitudinal follow-up.

And so rather than concentrating on the typical RCT that we see in most FDA submissions, at least for the Class III devices, I think you have to accept a high-quality observational trial and make sure that the specifications for what that trial consists of are obtaining the highest quality of evidence possible.

Probably more important than ethnicity in these groups is kind of like a genetic basis for their disease, they'd probably have a much bigger influence, frankly, I would think, and having that diversity of that may be important, or maybe targeted to one disease population such as Duchenne's or spinal muscular atrophy or something.

DR. LARSON: I think there's clearly a role for registries, and within the pediatric spine community there are several high-quality registries that are built upon manual data abstraction and many, many hardworking study coordinators and I think that's probably the future, but how do we open up those super high-quality registries to more patients and again, get broader numbers and not just a limited number of centers such as harm study group, pediatric spine study group, how do we broaden that to a general population?

DR. BRODKE: I was going to comment real quickly on a couple things that you brought up earlier, which I wanted to underscore again because I thought they're really important points. I think this issue that you brought up around economic diversity is a particular challenge and some of the points that Dr. Golish brought in around support from the industry sponsor that may be doing an IDE-type study, specifically to make sure we have the kind of diversity needed, is probably one of the answers to that issue, but it is a challenging issue, particularly in the areas out in the West, like ours, where travel is almost mandatory, like long-distance travel for particularly unusual problems, but even common

problems, we have patients traveling long, long distances for care. I also liked your idea of trying to find sites that may not have traditionally participated in IDE-type studies that may have the populations we seek. And so trying to specifically focus on a certain number of sites that are like that may also allow us to broaden our patient demographics.

DR. LARSON: Just like we do a risk severity assessment score for patients, right, someone who has diabetes and is overweight and has high blood pressure is going to have a higher risk of any type of procedure. The different sites are going to have a higher risk of loss to follow-up. So I mean, if we had kind of a site assessment of "this is a high-risk site but high reward" because we're going to get a different slice of society and then from a regulatory standpoint providing more funding to those, if you wish, more novice sites or less experienced sites, because we understand the patients that they are able to recruit are valuable.

DR. BRODKE: The real challenge, of course, is making sure we have the follow-up we need because that's going to be the biggest challenge in -- it's the biggest challenge we face anyway, but it's even a bigger challenge in sites that are inexperienced or in populations that have a hard time with follow-up.

DR. LARSON: For sure. And I think the virtual visit is really a key piece. I mean, I have been so pleased with the ability to kind of access patients remotely and they've been pleased, as well. So I don't know, is that an open topic for discussion? Can we do 2-year virtual follow-up for an FDA IDE study?

DR. ANDERSON: Yeah, I think that's a great question and I'd like to hear from the leaders of the FDA about that. Should that be built into the studies now? Is that acceptable to people? Obviously, you can't do a neurologic exam, but to me that's a little bit less important than the patient-reported outcomes, but I'm sure we'll talk about that later.

DR. LARSON: Well, you have your toe walk and heel walk and squat and I mean, you

can do an approximate --

DR. ANDERSON: A little bit.

DR. LARSON: -- you know. Yeah.

DR. DEVLIN: Dr. Anderson, I would say FDA is open to well-defined novel study designs, so I think certainly the virtual trials will become increasingly important in the future, as will registries which would help this area, as would methods of automatic data capture which would certainly take the burden off investigators if we can improve those systems.

DR. GOLISH: So this is Ray Golish. You know, that's a great comment, Dr. Devlin. And you don't have to be an FDA official like Dr. Devlin or Harner to remember that the FDA does have alternative regulatory pathways which are relevant here. Humanitarian device exemption is certainly one of them for the populations that can be defined and for which you can study safety but cannot, simply cannot execute on a full study of efficacy and that pathway has been successfully used. Similarly, the notion of a custom device, the definition of which is very narrow, five or fewer per year, but just 3-D printing or just image guidance does not make it custom, but for very limited populations and very specific diseases, those are relevant. And then reminding ourselves that many relevant devices are Class II devices and the sponsors wanting to undertake studies for particular populations outside of that and you need to be careful, but that also applies to Class III devices and IDE studies.

So in addition to real-world evidence, which is a great shining hope not only in the pediatric discussions that we have at the academy and NASS and with the Agency, but even outside of pediatrics for those underserved and underprivileged populations, that is the tool that we all are trying to work in the middle to build for all the reasons that have been described by all the panelists nicely.

DR. LARSON: It's hard to get industry interested in a big pediatric device trial, all the

limitations we've outlined, but at the end of the day it's just a very small patient population, so they can't recoup the costs of doing a high-quality, large prospective study with controls.

Dr. Brodke, go ahead.

DR. BRODKE: I'm sorry, I was just going to hit number -- the third question for a moment, just because I think it's a key piece, and that is we see different prevalences of disease in different populations and we may be bound by that issue in trying to build out our diverse demographic population to some extent. And we assume, I guess, maybe rightfully or maybe not, that the effect of a treatment on a given disease would be similar even though the prevalence of the disease is different in different race and ethnic populations, but it sure would be important to make sure that that assumption bears out. And so we may need to figure out how to focus and again, that's going to be to some extent a sub-design issue when we have this kind of disparity, I think.

DR. DEVLIN: Dr. Brodke, you may have seen the *New England Journal of Medicine* article this week where they are making it a requirement for the new studies going forward to list these types of factors and the prevalence of the disease in different populations and consider all of the factors we're talking to today. So I think this is going to become much more in the forefront as time goes on.

DR. KEBAISH: Another thing, and I don't know if that would be practical, but is requiring a certain number presenting, but the actual problem we're investigating in that population. So if you have it more commonly in an underrepresented group, you should at least have a reasonable representation, otherwise you'd expect the results not to be valid or at least as representative in real-life data. So something that may kind of allow us to or at least -- because we all want to follow the rules, so if that rule is imposed, you can't have a study that's designed to test a very rare disease that is rarely found in the majority of the population, we need to have representation. It may make it harder, but it may also

incentivize industry and those participating in the studies to recruit those patients.

DR. DEVLIN: For the small populations, though, I think we could use more of the humanitarian device exemption-type studies as was done in the spine arena with the tether devices. So there may be more work in that area in the future.

DR. LARSON: Dr. Devlin, could you expand a little bit more on the HDE process? I mean, it's been wonderful for pediatrics to have these two new devices approved and one thing I've appreciated with the HDE is that role of postmarket surveillance or at least post-approval studies being required as part of it. Do you have any comment? Do you think there will be more devices coming forth through that HDE pathway or is that a good pathway for peds?

DR. DEVLIN: Well, I think it can be very valuable because the standard is safety and probable benefit, so it doesn't need to have as large a study. So, because of the exemption from the effectiveness requirement, a smaller study which would lend itself well to a specific pediatric subpopulation would fit very well into that niche. And certainly, doing an HDE would not preclude a manufacturer or a group of surgeons from coming in with another marketing pathway in the future once they collected more robust evidence down the line.

DR. LARSON: Again, we have two devices approved by the FDA through the HDE pathway in pediatrics and both are requiring post-approval studies, which I think all of the surgeons have appreciated that there's ongoing data collection with these devices that's mandated.

DR. DEVLIN: I think it's far preferable to at least have an approved device rather than having experimentation with un-cleared or off-label use because then you would never really learn what the outcomes of those devices are because they're not systematically studied. So by having the postmarket follow-up, which in the case of the tether devices is

through a registry embedded in the post-approval study, that really addresses both aims and makes it easier for the investigators and makes it more scientific that we can actually collect the data on the device that's being used.

DR. LARSON: A large number of pediatric orthopedic devices are used off label, which is a pathway where we can't prospectively study it, so the HDE pathway is very valuable to be able to study and learn from the techniques at multiple centers at one time.

DR. HARNER: Thank you. We're going to -- that was excellent and we're going to move to our third and final topic that Dr. Moazzam is going to present.

So Dr. Moazzam, please, put your questions up there. Thank you.

DR. MOAZZAM: Good morning and welcome to the FDA. Thank you for your input on the following topics under the category of challenges with defining a target population and enrollment.

Question 1: How do we define a study population for a new technology when the target population is traditionally treated without surgery or use of a medical device?

Question 2: How do we balance enrollment criteria necessary to isolate a treatment effect with the need for so-called real-world patients?

And finally, how do we strike a balance between diversity and the myriad of challenges which can limit participation, for example, noncompliance, lack of follow-up, or comorbidities?

DR. BRODKE: Dr. Moazzam, maybe I'll start. I think these are all excellent questions and very difficult to answer, even more than the last set, and particularly number 3, striking a balance in diversity with the challenges posed, which we've spoken a little bit to already this morning. One of the solutions, to me, which I think Dr. Larson already actually brought up, was about trying to do this postmarket surveillance as a formal part of understanding how treatments and products used in the real word affect patients in the real world and

that's often not attained with IDE studies. You know, the diagnosis is very narrow, the treatment group is narrow so that we can actually identify a treatment effect or not, and we miss maybe in certain occasions the most frequent way something is used. And so having registries and postmarket surveillance, both of which allow us to look at, I think, real-world treatment, is really important and I think probably a tool that's underused today for -- at least for the FDA.

DR. ANDERSON: I'll comment on the first question, which is how to design -- if I understood it, how to design trials where the control group is non-operative treatment or the predominant treatment at least currently is non-operative treatment, and I've always said that defining the control group is one of the hardest aspects of designing a clinical trial. And the control group has to obviously represent, at least as far as we can tell demographically, disease severity, comorbidities, as the treatment group. But that oftentimes may not be the case and so consequently, you have disparate populations that you're studying.

For instance, I'm doing a trial comparing the use of an anabolic agent in osteoporotic patients undergoing total knee, but I could not do that randomized controlled trial because I cannot give osteoporotic patients a drug in the placebo group. So I couldn't do a placebo control trial, so now we're doing osteopenic patients who are not getting treated against osteoporotic patients who are. Obviously some deficiencies in that trial design.

But I think the control groups, the non-operative control groups, that's one reason why we don't see them much in these FDA submissions, is usually comparing against a predicate surgical technique like a fusion or a fusion with X implant that's been around forever to get rid of that. But I don't have a good answer other than trying to control as many factors as possible. Propensity matching in a randomized clinical trial, you know, maybe that's helpful to narrow -- at least try to get those patient populations as close as

possible.

DR. LARSON: I have our IDE looking at a non-fusion device for scoliosis at Mayo and we're trying to enroll fusion controls for scoliosis, but at the end of the day our fusion controls are going to be older than the non-fusion patients and we're trying to get them both prospective, but it's very hard to find the controls because everybody wants the new device. And in peds, as we heard, it's very hard to randomize, I think impossible in this instance.

DR. BRODKE: Yeah, this is really key, trying to identify the control group and one of the challenges and we've seen this many times in the past, is about the time the study is done years later the control group is really not even relevant anymore, like the whole treatment paradigm changed during the time of the study and that's an additional problem, is really identifying the control group that will be a lasting control.

Trying to enroll patients into a study where the control is non-surgery or non-use of a medical device because that's traditional, when they think that the new thing that they're getting if they stay out of the study, because then they could be treated with the new thing, is really -- especially if there's one on the market already, is really impossible to do. So once we have something similar on the market, trying to do a control that's non-operative is really difficult.

DR. ANDERSON: Yeah, I might add the crossover from control to treatment is pretty powerful in these trials, for instance, in the randomized controlled trials of vertebroplasty, 40% of patients in the control group eventually got vertebroplasty and yet, if you do intention-to-treat analysis, it looks -- the ones that got vertebroplasty or initially controls are not graded if they were controls. And so that type of crossover -- and it can go the other way, as well, but it's hard to undo a surgery, the crossover is almost always conservative to surgery, you have to build that in your statistical model, how you're going

to handle those patients and that would be critical for the study design.

DR. GOLISH: This is Ray Golish, I'd like to comment on the part of the question that mentions follow-up and diversity inclusiveness. You know, from the sponsor's point of view, a high rate of follow-up is absolutely critical and a loss to follow-up is absolutely imperiling no matter how much goodwill you have. The statistical tools for adjusting for loss to follow-up can help you squeak over the finish line, but keep in mind the number of having 90% follow-up, all data points, all time points at the 2-year final time point is really a starting point for a good trial. That is a very, very high bar for anybody who's run certainly an unsponsored trial, even a sponsored one.

So, because of that, in order to increase diversity inclusiveness, those investigators who have access to underprivileged and underrepresented populations are crucial, as Darrel Brodke was emphasizing, and then incenting those patients in order to participate by virtue of their personal relationship with that investigator and any financial incentive that could be increased, give to them to do that in a way that's compliant as far as the Agency is concerned, but that allows a sponsor to pursue that as an objective is really important.

And then I'd like to revisit something Darrel Brodke said about the real-world evidence and post-approval studies. You know, post-approval studies are a new normal, they're a great tool the Agency has, they are challenging for sponsors, so I'm not aware of any Class III device where path 1 has been -- has not been a 5-year follow-up of the 2-year results, and perhaps a path 2 being real-world conditions of use to capture that larger population and the difference between efficacy and effectiveness. And those are extremely important studies, but they are difficult to run and can be burdensome and acts as sort of a tax on the device's functioning as time goes on. That makes, from my point of view, real-world evidence and registry data even more important for those passes than they are for the possibility of pivotal trials to begin with, because they make the ongoing function of the

device sustainable as opposed to just getting over the initial regulatory finish line for approval.

DR. LARSON: And you end up with a negative feedback loop. Because it costs so much to do the IDE and then so much to do the post-approval study, now the cost of the device is very, very high and many insurers won't provide you access to the device. And so it's a big challenge and I think technology maybe is part of a solution.

Again, if you could imagine a world where people are enrolled in kind of low-cost, automated data capture from their healthcare system or electronic device-based data capture for patient-reported outcomes, some kind of lower-cost methodology for doing these post-approval studies where we're not depending upon study coordinators calling the families and doing paper forms, filling out your Oswestry Disability Index on paper and a study coordinator adding that up. So I think automation has hit many, many other sectors of our society, maybe more so we'll see automation bringing down the cost of these post-approval studies.

DR. KEBAISH: Yeah, that's a great point and one of the comments, I think the concern is we have most of those unusual problems that we investigate are basically studied at centers, say, that tend to track these patients and there is not necessarily where it's most prevalent and that obviously becomes a bigger challenge for the follow-up. And if we can automate those processes to some degree, we're able to still continue to have big centers that attract these patients from the underrepresented community that may not be in proximity to where the trials are being conducted, so we're able to continue to have the follow-up. Yeah, I think the biggest challenge is always going to be the follow-up and how do we keep that high-level follow-up that is required for some of the devices that are being tested.

DR. DEVLIN: This is Vincent Devlin, I agree with Dr. Kebaish. You know, with regard

to the follow-up challenges, our efforts in registries and other real-world evidence are really directed at addressing this, so I would just keep that in mind, I think that can help across all of the areas.

And then just a comment on the control groups, I would say that control groups in spine are very challenging because there's no consensus on what are the standard elements for non-operative treatment. And also, it's generally not a parallel treatment. Usually, people will fail non-operative treatment and go on to a surgical treatment. So to compare an operative with a non-operative control in which the control patients have already failed the non-operative treatment is not meaningful.

DR. ANDERSON: And you also have the problem of evaluating adverse events in a non-operative group versus the surgical group because obviously, there's a certain degree of surgical invasiveness that goes on, which it correlates strongly with your risk of adverse events, so is a that meaningful effect? Yeah, if you're going to do surgery, you're going to have more complications, probably.

DR. DEVLIN: Great points.

DR. ANDERSON: Hey Vince, I have just one comment about your presentation. The term SSI, that is really utilized throughout epidemiology as surgical site infections and I find it quite confusing, so I really urge the FDA to change that term.

DR. DEVLIN: We can emphasize that in great depth in a later discussion. We welcome all input.

DR. HARNER: Okay, we're going to end this panel, it has been an outstanding discussion that I'm sure, as we move forward, is going to be very valuable to changes that we look at with the FDA and clinically with our clinical partners and industry.

So the next session is on industry's perspectives and it will be presented by Ms. Janice Hogan, who has been involved in medical technology for over 25 years. From

her engineering training at MIT to work in the pharmaceutical industry to her current practice representing medical device companies before the FDA, Ms. Hogan has focused her career on the intersection of technology, regulation, and healthcare.

MS. HOGAN: Thank you very much, Dr. Harner. It's my pleasure to talk with everyone today. Fascinating discussion this morning and I'm happy to be a part of it.

One thing I will add maybe to the last discussion, for all the 25 years that I have been doing this in orthopedics we've tried many ways to improve the representation of underserved communities in studies and one idea I had listening to you all, I think technology and use of registry data are both really good methods, but companies have done exit interviews with people who opted out of clinical studies and one thing that's been cited over and over again is the time burden, especially in orthopedic studies in working populations.

We do somewhat better with other types of studies that are less burdensome, say in diagnostics, but we've often tried to add up the hours of time commitment for patients and many orthopedic studies, if there are seven or eight visits that each takes several hours, we cannot compensate people for all of that lost work time, childcare, elder care, all the things that might offset 20 hours of their time, we just can't give them compensation that would even that out for them ethically. And so it might be helpful, when new IDEs are coming in, to think about and add up the total hours of burden on the patient because we see that we do much better on this issue in diagnostics, in pharmaceutical studies where the time burden is less. Orthopedics is that tough convergence of patients who may still be working and quite a lot of serial visits that take quite a bit of time.

With that, I want to move on to a couple of topics that come up repeatedly in my practice with a lot of different companies pursuing spinal device studies. I think one of them probably makes more challenging the diversity issue that's been discussed this

morning, but the other is hopefully offsetting.

So one thing I talk to clients about all the time is the use of foreign clinical data in spinal and other orthopedic device submissions. As most all of you probably are aware, in Europe there has been a change in the regulatory system such that it's going to become more and more common that companies will need larger, I'll say more FDA-like clinical studies before they can obtain CE mark. And since that is different than in the past, it adds to the incentive to try and run multinational studies in the United States and Europe or elsewhere, for example, to try to meet the regulatory needs of two different regulatory entities, the FDA and the foreign regulator.

Whenever we talk to clients about using foreign clinical data, we say certainly it's possible, we don't discourage it, but you have to think from the beginning about whether you can enroll a population that is representative of the United States, depending on where you're doing the study. And we've looked very closely at many centers in Europe and in fact, you know, many times it's much more economical to run part of a study outside the United States, the overhead is less, but we have trouble with it because we cannot identify centers very well that will recruit a U.S. representative population. Not just in terms of race and ethnicity, but even of obesity and other factors.

And so while we would like to move more in this direction, it does raise even more concerns about diversity and representation. And so that's an issue that's common to a lot of my clients that we're trying to work on, identifying centers outside the United States that can deliver a representative population.

When we look into this more deeply, we know what we have to do to meet FDA's requirements. We know the study, if it's done outside the United States in whole or in part, we know it's possible. We have to do something that's compliant with good clinical practice and we have to prove our case, at the end of the study, that the population in Europe is

representative of the U.S. So if, for example, we're doing a study 50% in Europe, 50% in the U.S., at the end of the day we look at what we get and we hope that it will all be poolable. What happens if it is not just a risk factor that companies carry. And sometimes we look very deeply at the data and we may or may not find a satisfactory explanation if the data turns out not to be poolable.

So I would expect that for the future this is going to continue to be an important topic because of the changes in the regulatory system in Europe and that an increasing number of companies will try to run multinational studies for the reasons that I said, but we're going to have to continue to watch and try to manage these challenges with representation and poolability.

FDA has been good in working with us on approaches to labeling if, for example, we see some differences between U.S. and foreign cohorts. A couple of years ago I did a review of the last 100 PMAs that have been approved, not just in orthopedics, and I looked at how many of them included foreign clinical data and about 40% of them did. And there was no evidence that it made the PMA approval process more difficult or longer, so certainly it can be done. There are a number of orthopedic PMAs that have been approved based on foreign clinical data, but for all the reasons I mentioned, we advise clients generally to stick to sort of a 50/50 ratio to give us some balance between U.S. and foreign data.

The other topic I wanted to hit on just briefly is the use of real-world evidence. As has been mentioned a few times, what do I mean by real-world evidence? This can be anything from registry data collected outside of a formal IDE study in the U.S. or abroad, or even more real-world than that, health claims data with the rich datasets that are available now through electronic health records. We are really starting to develop more and more innovative ways of tapping into those data sources. A whole cottage industry of vendors

has come into existence that help us to mine electronic health records in a way that's compliant from a privacy perspective.

And I think that this might be another answer in addition to the use of technology, because one thing we've seen, for example, in one project I'm working on we are using health claims and we have a hundred thousand records. Most companies that have done studies, prospective studies, have 10,000 patients. We don't have all the depth of data of a 10,000-patient study, but on the other hand we have a hundred thousand records with really good diversity.

So maybe there is a balance in between where the FDA could, for example, consider some classic randomized or nonrandomized prospective studies, but to help to enrich our data with better population diversity, claims data or other types of real-world evidence may contribute to the answer. And the FDA has published a report that talks about the ways that real-world evidence can be used and there are a lot of successes. So far, most of the successes we see are for device modifications, for labeling expansion, not for the original approval, but hopefully that will continue to grow moving forward.

One other comment I would make is with COVID, we've learned a lot of things that we didn't have to do before to keep clinical studies operational and I will say that the FDA has been great about this, the FDA has really helped sponsors I've worked with to be flexible, use alternative methods, use virtual visits to keep studies running. We have also learned that geographic dispersion of sites helps us and having multinational studies actually has helped a number of companies to keep going. When things were very restricted in one geography, at least they could continue to enroll in another. Of course, we do none of these flexible options without keeping a strong eye on data quality and patient safety, but I do echo the comment that Dr. Larson made, I think, that I believe technology and virtual visits will contribute to us improving the diversity and representation in clinical

studies. Thank you.

DR. HARNER: Thank you, Ms. Hogan.

We're now going to start the last session Q&A before lunch. This session will be conducted by Dr. Vincent Devlin, who will be moderating the Q&A in addition to Dr. Ronald Jean, Director of the Division of Spinal Devices in OHT 6. Dr. Jean has served with the FDA for over 16 years, both as a reviewer and various leadership positions.

To the audience, please remember to send your comments and questions to the OHT 6 e-mail that is on your screen.

Dr. Devlin and Dr. Jean.

DR. DEVLIN: Thank you. I believe Dr. Jean has the first question.

DR. JEAN: Yeah, good morning, everybody. I know we have a whole afternoon left of clinical discussions and we've already had some incredible dialogue already, so this is really great.

Polling from our OHT 6 feedback box, I will sort of pose the first question to our clinical panel and that is in your experience, how has the COVID-19 public health emergency influenced conduct and planning of spinal clinical studies at your institution, and are there any lessons learned that you would like to share which could be applied to optimize the design and efficiency of future spinal device clinical studies?

DR. LARSON: I think all of us felt the COVID pandemic in different ways. I'd say the biggest challenge at our center has been just the tumultuous time at the Institutional Review Board, more limited resources as far as our research infrastructure, many of our study coordinators got reassigned or furloughed during the height of the pandemic but again, valuable lessons, as well, as we heard from Ms. Hogan, as far as virtual visits and learning to work differently that I think will affect all of our careers from this day forward. More Zooms, right?

DR. ANDERSON: Yeah, I agree, it totally changed our ability to enroll patients, we were stopped for 6 months from even enrolling a single patient, and so it ruined those ongoing studies that involved consecutive enrollment of patients. Yeah, hopefully the editors, when they review the article, won't punish too much for that. We also had interventions, we couldn't apply diagnostic tests like DEXA scans, we couldn't obtain them because DEXA was closed down at our institution, so it really ruined ongoing studies during that time.

However, we did learn from that and particularly using a lot more telemedicine built into the follow-ups, we found that to be actually quite useful and are building that into the protocols.

We also had a disruption between the time of enrollment and when surgery was occurring. For our study, that was particularly important, we had to have correct timing and with all of the starting and stopping elective surgery, we had people ready to go to surgery who got their pretreatment and then surgery didn't happen. And then other ones who were enrolled in the study who needed the pretreatment got called up, we can do your knee next week and they did good. So it had huge changes and fluxes in our study protocols and we have so many violations of the study protocol that we're going to have to explain, but we're hoping that the mask of COVID will just paste over any deficiencies. But that's what happens when you have an unanticipated event.

And I think from the FDA perspective, you also have to realize that outcomes of ongoing studies done during this period are going to have a lot more variability just from the psychological effect of what's gone on, you know, people -- I have to believe depression, for instance, which is obviously a real critical component to outcomes in spine surgery, is going to go way up just like the same thing happened after the World Trade Center, you know, the mental health of people are lot different and that may be reflected in

the clinical outcomes of these trials.

DR. JEAN: Very good points, thank you.

Dr. Devlin, would you like to pull a question from our feedback box?

DR. DEVLIN: Sure. Just one last comment on the last topic. My hope would be that the lesson learned is that there could be parsimony or more efficiency and limitation of the data collected and that we can be more efficient going forward because we can see what we really need in terms of data based on what we could get away with without doing based on the lessons from the COVID-19 pandemic.

DR. KEBAISH: And I'm hoping that, you know, at least one positive is that we've learned a lot. We shut down all studies for at least 4 months with the pandemic, but we went on changing protocols after that. So I think we'll have to build it in for a future study that there could be another pandemic that would stop us and how do we manage. So I don't think we'll ever have to completely stop even if we get into similar circumstances in the future.

DR. DEVLIN: Great point. Moving on to the next question, and this is directed to all of our panelists. Based on your experiences as clinical study investigators, what suggestions do you have for device manufacturers who sponsor spinal clinical trials regarding specific strategies they could use to optimize the design and efficiency of their future spinal device clinical studies?

DR. GOLISH: So this is Ray Golish. That's a really broad question, which makes it a good one, but I'll select a portion of it which echoes the previous conversation we just had. I'd encourage sponsors to think carefully about their site of service when you are running a hospital in a pandemic and focused on not only caring for the COVID ill, but the seriously medical ill, as well as doing surgery that is critical. Let's call it tumor surgery, let's call it vascular extremities. That remains a focus and a resource-constrained environment. For

many of the things being studied, even for Class III devices, alternative sites of services are available, whether that be all the way to an ASC that can sometimes be done faithfully and ethically, or to just a lesser site of service, those are things worth considering.

And then the final comment, I think, is to embrace what Dr. Devlin just said, some of the learnings have been around the efficiency of trials, what is really needed, final follow-up is needed and some interim points are needed. But if we can consider how valuable really are the data at time zero, 6 weeks, 3 months, 6 months, 12 months, 18 months, 24 months, up to that, you know, there's some room in there for innovation, I hope.

DR. LARSON: I'd say the trials I've been involved with, I'm always trying to get standard of care to converge as much as possible and avoid additional tests and additional site visits as far as trying to make the study protocol near what we're doing in real life, as closely as possible, both for the families and then also for the generalized applicability of the procedure once it is approved.

DR. ANDERSON: Yeah, I agree, I think both the previous two speakers are maybe implying take less data and make it much more clinically relevant data and particularly, the time points. And we don't need 57 different outcome scales, we need two or three targeted ones. We don't need flex extension and AP lateral X-rays at every time point or every visit and CAT scans at multiple time points. We need very targeted time points and a much more limited dataset. In my opinion, if there's too much data, it is actually a disservice. And I've found simpler studies "keep it simple, stupid," is a better approach to get a better answer.

DR. BRODKE: I would wholeheartedly agree with what Paul just said. I think it's incumbent upon us to really be much more targeted and figure out how to find out what we need to find out in as quick way as we can with the appropriate data collected at the right time. And I think sometimes we haven't done that, we've collected everything at every

time point and that's maybe not necessary. We'll talk a bit more, I guess, this afternoon about patient-reported outcomes but even with that, the cost of too much data is immense and we could do better.

DR. ANDERSON: Yeah, and I think it's also confusing at times because you have to account for it statistically and there may be a statistical anomaly there.

DR. DEVLIN: If there are no other comments, would you like to take the next question, Ronald, identify the next question?

DR. JEAN: Sure. So in looking at the inbox, there's a regulatory question so we can give our clinical panelists a small break. The question is how does an anticipated de novo classification affect the patient population selection? Can we start with a specific target population for the trial and expand later upon the first clearance?

So this question is asking about the de novo request process that we have and to understand it, you know, what we do under de novo is we essentially grant a request, so it's both an automatic Class III to Class II reclassification as well as a market authorization. And so once we create a classification for that product, then subsequent products that fall under that bucket can be cleared through our 510(k) substantial equivalence process.

So you know, like many questions, it really depends on there is a possibility that you could start with one population, collect the data, use that to support your granting decision and then even potentially expand that through 510(k), but it's all going to depend on how far you stray from the original population. It's also going to depend upon sort of what the -- how well defined the classification is that results from that granting of the de novo request. But one thing that -- you know, this is advice that I always give and I know that medical device industry veterans always give this, as well, is that it's really a good practice to engage FDA early and you can do that through a Q-submission, you can have a conversation about what your idea is, what your target population is. I'd also like to highlight that in the last

few years we have what's called a breakthrough device designation program as well as a safer technologies program or SteP, and both of those basically allow you to talk about your device at a more preliminary stage and if you are allowed entry into either one of these programs, you can have more frequent discussions with FDA or faster discussions on select topics. So I definitely would point the audience to looking through those materials, as well.

But again, I think that that really is something that can be unnerving for a small startup to approach FDA when they don't have all the pieces together, but we appreciate the opportunity to be involved in that early development stage and to give you the best feedback that we can at the time with the existing regulations.

Dr. Devlin, would you like to pick another question from our inbox?

DR. DEVLIN: Sure. We have a number of questions for the panel regarding any insights they would have comparing and contrasting the challenges between clinical trials in other countries with United States clinical trials and are there any strategies that they've seen or experienced that they would avoid, leverage, or adopt from their experiences with the international scene.

DR. ANDERSON: I have a comment and I thought about that, having been involved with the disc arthroplasty trial which was started in Europe and so we followed that data before we started the trial in the United States and I looked at the European results, which was nonrandomized, but if you look at the treatment effect of the disc arthroplasty group between Europe and the United States, it was almost identical. So I was very comfortable with the European data and I think that if people are going to submit that type of data, they should submit an analysis with at least that disease state showing the results for other treatments of that disease in Europe or Asia have correlated to what we've observed in the United States, so at least there is some reassurance that the outcomes are the same.

DR. DEVLIN: So for the cervical disc arthroplasty studies you were involved with, the

comparator would be anterior cervical decompression and fusion, then.

DR. ANDERSON: Yeah. Yeah, and like I said, but when I looked at the European results of the disc, the improvement on the Neck Disability Index and pain was identical to the randomized controlled trial in the United States of just the treatment group. So I learned, at least for cervical spine surgery, anterior cervical spine surgery, Europe and the U.S. are pretty equivalent to the best of my -- I didn't do a statistical analysis, but maybe somebody could do a nice meta-analysis there and compare the results among countries, and I've done that for some other diseases and in that case there were variations in outcomes depending on your country of origin. So you need to determine that if you're going to use outside the U.S. data.

DR. DEVLIN: That's very helpful. I know Dr. Kebaish has an international practice and I don't know if he would have any thoughts or insights on this area due to his extensive experience.

DR. KEBAISH: Yeah, certainly. And I think Dr. Anderson brought up really good points and we can rely on some of the high-quality studies that were done, at least with our clinical practice and we do like, for example, patients with TB. I mean, we really don't have enough patients with TB, I've treated a few patients, most of them are international. So there are some diseases that we still see here but really don't have enough population to draw any conclusion or conduct any effective studies. So I think relying on high-quality data from other countries is something that should be considered for some of the rare diseases that we're never going to really be able to study in this country.

DR. GOLISH: So this is Ray Golish, I'd like to springboard a little bit on what Dr. Kebaish and Dr. Anderson said, which is that, for me, consuming trial results is about having investigators in jurisdictions where people's practice patterns sort of match up and that's especially true when you're seeking, as Janice Hogan pointed out, multi-jurisdictional

clearances in the U.S. and OUS. The FDA has, in fact, approved Class III devices not only with OUS data, but with exclusively OUS data, I believe, because those indications did match up.

But you really need investigators who are basically thinking and feeling the same thing about their indications and that's true even in a pivotal IDE, and once you have a pivotal IDE, you've got your inclusion/exclusion criteria, you've got a hypercube, and the investigators are either behaving and putting their patients in that or they're falling out. But even then, you need to be reassured that those data are going to be extra-poolable, the difference between efficacy and effectiveness, when you finally do get that approval and you need to be reassured that the mentality of the investigators matches the clinical reality going forward.

DR. BRODKE: I kind of wonder if there isn't some financial pressure on sponsors today particularly with IDE-style studies that require CE mark and FDA approval that were running two simultaneous or two separate groups of studies that essentially doubled the expense and could that -- could there be some sort of common scale process where we have high-quality data collected transnationally. And I know there are complexities entering into that process, but it does feel like there is a lot of duplication and reduplication of effort in today's world where the data from many, many countries is an extreme high quality, unlike in past decades. And so I'd be curious to see sort of the corollary of that question, how people feel about that.

DR. ANDERSON: Well, I think you make a relevant point. Certainly you've been involved in international studies through the AO knowledge forum. It's seemingly relatively easy to carry out across multiple countries, many third-world as first-world countries, so it is feasibly possible.

I think like the arthroplasty trials all got started in Europe because they got a CE

mark relatively easily, whereas you couldn't start the IDE exemption through the FDA for quite a bit of time, and so they could at least get that clinical data going in Europe with the CE mark and that's why they started there. And I think that situation is no longer the same as -- and maybe the European FDA, whatever their equivalent is with our FDA, can get together and see about how to construct multinational randomized clinical trials so they can be going on concurrently in multiple countries. And you could include Asia-Pacific, as well.

DR. DEVLIN: All right, if there are no other comments, would Dr. Jean like to take the next question?

DR. JEAN: Sure, I'd be happy to. So looking in the inbox, the last question that I'm seeing, it's a specific question about a particular study that's registered on clinicaltrials.gov and the first part of that is asking is an IDE necessary, and we don't disclose whether or not the application is approved or not unless it's in the public domain.

But I would draw -- for the person who has that question and others that may be interested, what you can do is look at a resource we have that talks about significant risk versus non-significant risk studies and for clinical trials, if you have a significant risk study, you require an IDE and you're required to actually follow the full set of our regulations related to clinical trials. But if you have a non-significant risk study, you're only bound to follow sort of the abbreviated form. So that would be my response to that part of the question.

But the second part of the question asks what biocompatibility studies are needed to be a part of the submission. And so when we look at biocompatibility, we consider a number of factors and we have a guidance that's called the guidance on the ISO 10993-1 standard, so you can look that up to sort of understand our thinking and how we consider evaluation of devices from the type of tissue that is contacted, the length of exposure, and

there are other factors.

But I would also draw everybody's attention to a new tool that CDRH has. If you search for the Biocompatibility Assessment Resource Center, there's an entire sort of landing page on the web that will actually walk you through how we approach biocompatibility review for medical devices. So that's an excellent resource that really walks people through beyond the ISO 10993 guidance and definitely would encourage everyone to take a look at that because I will say, you know, the topic of today's workshop aside, biocompatibility has definitely been something that has been more challenging in the last few years since we had an updated guidance related to that area of review and we have many questions and we've devoted many resources to that. So I hope that both the question contributor as well as the general audience finds that information helpful.

And I believe we're both out of time and it looks like that was all the questions that we have, so I'll turn things back over to Dr. Harner. Thank you.

DR. HARNER: Thank you to all our panelists, to Dr. Devlin and Dr. Jean, for an excellent Q&A session.

We will now break for lunch. Let's plan for about 50 minutes, so we will -- sorry, we will resume our next session at 12:45 and that will be started off by Colin O'Neill, who will do an audience survey. So please plan to check back in at 12:45 and please send your comments and questions into the OHT 6-Feedback e-mail so we can collate those and present them.

Thank you very much, have a great lunch and thank you for an outstanding morning to all of the faculty who participated. Thank you.

(Whereupon, at 11:56 a.m. a lunch recess was taken.)

AFTERNOON SESSION

(12:45 p.m.)

DR. HARNER: Welcome back. We will open the afternoon session with an audience survey. Leading us through this will be Mr. Colin O'Neill, who is an assistant director in the division of spine devices within OHT 6. He has specific expertise in orthopedic device evaluation, regulatory science, clinical trial design, project management, and Orthopedic SMART device policy.

Colin, please proceed.

MR. O'NEILL: Dr. Harner. And welcome back, I hope everybody had a great lunch. To reengage everybody back to the workshop, we have a questionnaire for the audience and participants. On the screen is a way to access the questions and the Menti questionnaire. You can use the QR code via your phone camera or copy and paste the URL on the slide to participate. And there is a delay from the live recording of the Yorkcast, so I apologize, we will not to be able to comment live on your responses.

But so let's start with the first question, it's an easy one, and the next several questions will require some additional thought and I'll allow some more response time for those.

So the first question: Are you or your employer planning to conduct a clinical study related to orthopedic devices within (a) the next year, (b) the next 2 to 3 years, or (c) none planned?

(Pause.)

MR. O'NEILL: Looks like the majority of the responses are within the next year, and that's great. All right. Well, let's move to the next question.

The next question is: Cost is a known barrier to clinical studies. Aside from cost, what are the most significant challenges posed by orthopedic devices in clinical trials? This

is an audience word response and you get up to three responses per responder. For example, these challenges could include difficulties in enrollment, lack of follow-up, adverse event collection/categorization, lack of blinding, just to give you a few ideas.

(Pause.)

MR. O'NEILL: I see contracting is an answer. Enrollment and follow-up.

(Pause.)

MR. O'NEILL: Follow-up.

(Pause.)

MR. O'NEILL: It's a match between IDE requirements and clinical practice and spine surgery. Lost to follow-up in post-approval studies. Randomization concerns from subjects. Great. Thanks so much for these thoughtful responses.

If we can go to the third question and that is: What areas in spinal systems IDE and orthopedic clinical presentation FDA guidances would you like to see covered in more detail? Again, this is an audience word response with up to three responses per responder. And these guidances were published in 2000 and 2004, respectively, and provide important suggestions and recommendations for study methodologies that could help produce and present real-world scientifically valid data. And some of the information provided in these guidances includes considerations for selecting the right control group, inclusion/exclusion criteria, study duration, safety and effectiveness endpoints, and statistical analysis, as well as data presentation related to patient -- adverse event categorization, and patient success results.

(Pause.)

MR. O'NEILL: So K-I-S-S protocol design. Using definition. Postmarket surveillance, examples of real-world evidence. Alternate study designs, real-world data. Remote or distance electronic PROs that are automated. Consistent outcome measures.

(Pause.)

MR. O'NEILL: Endpoints. Endpoints for software as medical devices. Great, these are great answers and comments, appreciate it. Okay, I think we can go to Question Number 4.

Question Number 4 is: OHT 6 has met with key orthopedic registry stakeholders. What are some of the challenges that come to mind related to registry capture of data? Again, this is an audience word response with up to three responses per responder. For example, some challenges may include difficulties with data management, interoperability of data capture methods, incentivizing participation, utilizing unique device identifiers, just to name a few.

(Pause.)

MR. O'NEILL: Device identification, participation, postmarket surveillance data, follow-up, compensating for patients, compensating patients for good data and follow-up, more incentivization. Harmonizing protocols. Resources for data entry. EMR access data transfer.

(Pause.)

MR. O'NEILL: Data management and follow-up, that really seems to be a theme here.

(Pause.)

MR. O'NEILL: Adequate outcome parameters.

(Pause.)

MR. O'NEILL: Harmonizing data collection with international regulatory bodies where possible. These are great comments, really appreciate it. I think we can move on to the final and fifth question.

So the question is: OHT 6 has now held workshops on premarket, postmarket,

clinical review, and infections. What topics would be of interest in future virtual workshops? Again, this is an audience word response with up to three responses per responder. And for example, workshops can be held within premarket or postmarket review such as mechanical bench testing, animal testing, cadaver testing, additive manufacturing considerations, SMART orthopedic devices, recalls, post-approval studies, medical device reporting, just to name a few.

(Pause.)

MR. O'NEILL: Non-significant risk device determination. Enabling technologies, SMART devices, software as medical devices. Using real-world data for indication expansion, medical device reporting.

(Pause.)

MR. O'NEILL: Great, appreciate everybody's comments, they're really very valuable. I appreciate everybody's participation and I'd like to turn it back over to Dr. Harner. Thank you very much.

DR. HARNER: Thank you, Mr. O'Neill.

Our second clinical discussion is related to the evaluation for safety for spinal devices. We are joined by Dr. Anderson, Brodke, Golish, Kebaish, and Dr. Larson. In addition, Dr. Alander. And to Dr. Alander we will have Dr. Barton Sachs introduce our discussion topics. Dr. Sachs is one of our medical officers in the Division of Spinal Devices. Prior to joining the FDA, he was an accomplished healthcare executive, surgeon, educator, and scientific researcher, and has worked in both academic and private practice settings. For our first topic, Dr. Sachs will conduct the questions.

Dr. Sachs.

DR. SACHS: So thank you, Dr. Harner.

At this point we'd like to open the topic to the panel regarding the challenges in

capturing, classifying, reporting adverse events in spinal device clinical studies and will device registries help. With that, we ask the panel the questions: What are the best practices for capturing, reporting, and classifying the adverse events in spinal device studies, and what do you perceive as the role of registries in enhancing spinal device safety? Thank you.

DR. DEVLIN: If Dr. Anderson would like to lead off.

DR. ANDERSON: Yeah, thank you. I made a few prepared comments about adverse events. I have been heavily involved in research on adverse events, reporting adverse events, using the FDA SSED reports to try to report adverse events, so I have a fair amount of experience from looking at the regulatory documents on how these are handled and I'm going to make some observations.

Obviously, adverse events are essential to evaluate the safety of a medical device and the FDA has a definition which I think is very usable, an adverse event is any undesirable experience associated with the use of a medical product in a patient. And the collection and counting up of adverse events is a balancing act between the regulatory needs where you need a very broad definition because you're looking for those weird kind of side effects that we would not anticipate versus the medical and surgical view of them.

From my perspective as a surgeon, most of the adverse events that are reported in the SSED reports, for instance, are not helpful at all. In fact, they're confusing to surgeons and so we have a much narrower list that are much easily relatable to the device. So there is a distinction when you look at the regulatory issues, but when you compare that to what is needed by the medical team. And it is important to realize that although the FDA is to regulate in the regulation field, you do want that information to be usable by the surgeons so we can make the best judgments for our patients and also our patients can be an aid in that decision.

Important things that are attributes of adverse events, most important is frequency, but equally important is severity and it's very difficult if somebody's using uniform definitions of severity. There's also causality and attribution, was this related to the procedure, to the surgery, to the disease and a variety of other things or was this directly related to the medical device itself.

You also have to evaluate these at time points. The early ones are maybe more surgical related, they're operation related, whereas longer ones may be more device related.

Then the etiology, what caused this event, and these are always guesses and they need to be arbitrated, but Sohail Mirza has an excellent review of this in *BMC Musculoskeletal* and he used the Harvard medical protocol where there are 18 etiologic types and showed in spine surgery populations that you can have good agreement among surgeons under what was the type or the cause of this.

You also have to identify risk factors, there's certainly -- and we talked about this this morning, is certain demographic groups, genetic groups, ethnic groups, have higher risks for complications and that needs to be accounted for somehow.

And then also there's surgical invasiveness. We talked earlier about conservative non-operative treatment versus an intervention. Well, obviously they're going to be radically different complication profiles. And another example would be an interspinous spacer placed percutaneous on local compared to a decompression infusion. Those are real different operations and you have to anticipate different complications.

Statistically, I don't think most studies are really powered to show differences or non-inferiority of safety, they are powered usually for clinical outcomes but not so much for the safety and even if you do, what is the primary outcome variable, which of those complications is the most important in your analysis and what time points do you really

want to analyze this at for safety? And how do you account for severity? For instance, a urinary tract infection is a complication, it may require antibiotics, but a spinal cord injury is a devastating complication and do you count them the same or do you have to have some accounting for severity?

And just to identify some of the challenges is we're collecting too many adverse events, many of them are totally unrelated. For instance, in the BRYAN cervical disc trial, which I reported, a patient got hit with a golf ball and that was an adverse event, that doesn't strike me as a medical event.

There's also clustering of adverse events that go on, is that one adverse event leads to many others. For instance, a retropharyngeal hematoma is going to cause dysphasia, hoarseness, airway obstruction, and revision surgery. Is that one complication or is that five complications? But it's typically classified as five complications.

There are inadequate and variable different definitions of adverse events among investigations. The revision surgery definitions, and I reported multiple publications using that, and I still don't understand that and I still struggle trying to classify that. Now it's called SSSI, SSI, which somebody will have to explain to me what that is.

And another one, the most common complication of disc arthroplasties was pain episodes. Well, to me, pain is an outcome, not necessarily an adverse event, and so adding those into the adverse events made the totals go way up and make it look like oh my God, this was a horrible operation when, in fact, a patient just called in because they were in a trial with pain, which if they weren't in the trial they may have never even called in and may have gone away in 24 hours.

So recommendations. I think we need to get a shorter list of the key adverse events, standardize a severity classification such as the BINGO, there's also something called SAVES which has been published. Define a priori the variables that are going to be included in the

statistical outcomes as well as the time points. Have a uniform definition library that everybody in an FDA trial could utilize. Statistically figure out how do you deal with these clustering of events where a patient gets multiple compounded ones, and eliminate those that are really related to outcome or other events that happens in patients' lives. And again, I totally agree that long-term results need registries. The problem with registries is that there just is not enough data going in to be helpful at this point in spine compared to the arthroplasty, which are more robust now. Thank you.

DR. LARSON: A lot of great comments. I know within the peds world we've used this Clavien-Dindo-Sink classification to look at the severity of complications and as you said, a spinal cord injury, something that's permanent and it changes every day of your life thus forth is different than a superficial wound infection, and finding maybe a severity rating system and not spending 90% of your effort on the very minor complications and instead trying to focus on the very severe ones that we are certain are devise related would be helpful.

DR. KEBAISH: Yeah, I echo the same comments. I think the multitude and diversity of adverse events we collect makes the -- you know, any study very burdensome and how do we maybe streamline those adverse events for making -- maybe we don't have to focus much at all on UTI after someone had a disc arthroplasty or something completely unrelated or they got hit by a golf ball, but how do we walk a fine line between collecting too much and maybe collecting too little? But I think that's probably the biggest driver of cost of some of these studies and also time consumption.

DR. LARSON: We struggle with categorizing pain episodes. Many people, after having back surgery, have episodes of back pain and Dr. Anderson mentioned this as well, that I have a lot of reported pain episodes on my IDE and how do you know that's again device related or different than the natural history of scoliosis or a spinal deformity?

DR. BRODKE: I'll maybe hit the second question a bit on registries. Registries, it's an interesting topic because registries can be really deceiving and difficult and problematic and they can be really quite good if done well. There aren't a lot of examples of well-done registries, but a few that we could look at. Clearly, the problem with registries is loss of data or cherry-picking or not having the kind of data needed to really assess the outcomes over time.

A good example of where we should be heading with this, I think, is what -- and Paul mentioned this already, is where the joint registry is set at AJRR. With automated data feeds from hospital, don't require the surgeon team or the research team to input data and the ability to follow implants and complications that relate to ER admits, readmission to the hospital and revision surgery are all on an automated process that can really enhance the reporting and across state lines, as well, by the way.

American Spine Registry, ASR, is just getting off the ground with the hopes of paralleling the effort that AJRR has done and hopefully, as this kind of rolls out and if it answers some of the promises that it seems to be trying to answer, this may be a really good way to manage the number of things we're talking about in longer-term follow-up.

DR. ANDERSON: One of the limitations of the registries, as far as adverse events, is hopefully you will be able to identify explants of devices/revision surgeries with that, but identifying this long list of adverse events that occur regularly is not feasible and nobody has shown me that using simple codes to identify those is as effective as having research nurses. So that's going to be a limitation, is the amount of granularity of the data from registries to look at adverse events other than revision surgeries, explantations, maybe -- even surgical site infections, for instance, is quite difficult. If you use the CDC methods, you really need an infectious control nurse to go through each chart to confirm that, so that is very difficult to do in a registry.

DR. LARSON: I'd argue potentially a patient-facing registry or direct-to-patient registry may have the ability because patients often know whether it was the same piece of their spine revised or whether they were in a car wreck or again, this kind of crowd-sourced system where we're asking individuals to monitor their own health data for the report back to the registry.

You said the NSQIP registries, you know, high-quality registries need kind of on-site nursing review to generate that high quality of data. Can we skip the healthcare provider and go straight to the patient? There's privacy issues, certainly, with that type of approach.

DR. GOLISH: So this is Ray Golish. I'd like to layer on to some of the thoughts that were already offered around granularity, the work and intensivity that requires to maintain a database.

You know, there's something, some kind of sweet spot in the middle. I think the success, early success and then later on near failure but then ultimate success with NSQIP is illustrative of that. It's a little too full, but it's nice, in lots of people's eyes, to have enough data to be hypothesis generating, but it's much more important to have clean, clear data for those hypotheses that you can't articulate up front and the best example of this is closed claims data, it's the ultimate source with a big *n*, meaning capital envy, but not only a huge amount of classification data but huge amounts of missing data. And the real problem with this is that it really has the potential to be a Type 2 error machine.

The effect size and the power therefore of any data-based hypothesis influenced by three variables, you know, the difference between two groups and the variability each of those two groups. Heterogeneity and classification noise of the database increases that variability and therefore dramatically decreases your effect size and power. And so if you like to prove that things don't work, and there are plenty of people who wish to do that with bias, it's a good thing to cull through some database like a closed claims database with

a very large *n*, but very poor effect size is because of heterogeneity, because of ineffectiveness. And so we have to be really careful with that and deciding what hypotheses we're going to test, and can the database support that regardless of whether it's an IDE, a randomized controlled trial, a closed claims database or hopefully, some registry data from now on. I'm very optimistic about the power of registries, I vigorously applaud, as do many people, FDA's openness to registry and real-world data. The question is what question are you trying to answer and that needs to be thought of early.

DR. BRODKE: To layer on the challenge of large numbers, hence hundreds of thousands of patients, for example, in very large databases is something else in addition to whatever goal you're setting and that is that you can find significant differences between groups that are meaningless in terms of real clinical meaning and that's -- you know, really understanding the significant differences and clinical differences are really a different thing, and identifying very clearly what the clinical differences are is really important, particularly in relation to some of these outcome studies that we'll be talking about.

DR. ANDERSON: Yeah. Another word of caution about databases. If the database was created as a quality improvement project, then it is very difficult to use it to answer research questions. For instance, the Own the Bone database for secondary fracture intervention is very good at keeping track of what the patients look like when they get entered, but you can't really answer any hypothesis-driven questions with that. NSQIP is kind of like that in that it's a quality improvement program, you measure your risks against benchmarks and so you can see where you're deficient and hopefully put improvement programs in, but to attempt to do research out of that becomes very problematic.

(Pause.)

DR. HARNER: If there are no further comments or input, we will move to our second topic from Dr. Alander, he'll present this.

So Dr. Alander.

DR. ALANDER: Yes, thank you, Dr. Harner.

The second topic for discussion is centered around the challenges in reporting of subsequent surgical procedures in spinal device clinical studies. Questions for the discussion are: What different types of surgical interventions follow index procedure and how are these interventions best classified?

The second question is: How do subsequent surgical interventions impact patient outcomes?

DR. LARSON: I was going to speak a little bit to this as a pediatric orthopedic surgeon. We have a variety of very complex patients whose spine is still growing and kind of our baseline test case was traditional growing rods where we would implant the device and then go back surgically twice a year for 5 to 9 years on a planned approach to grow the spine for the child. And we have new devices now available that don't require that surgical lengthening, but all of these devices have a very high failure rate. So again, with the traditional growing rods, we're looking at a 50% planned reoperation rate within 2 years. So trying to figure out what those lengthening procedures are and what are the unplanned device failures has been very challenging for our field.

In addition, with these kids that were going back to the OR twice a year for 5 to 7 years, we would often delay the tune-ups, you know, so we'd have a piece that would break or fail or become dislodged and we'd wait until that routine lengthening procedure to fix that piece and now we have devices that can be lengthened magnetically or don't require lengthening but still have need for, you could say, these tune-up or revision surgeries where something has dislodged or broken or is no longer functioning and we require a return to OR. So certainly, for the under age 10 population it's known that there's going to be a return to OR and many of these procedures are set out as two- or three- or four-stage

adventures in care and we warn the families of such. So with that context now kind of looking at the newer devices coming out for children who are over age 10, it becomes a very important question, what's the comparison group? Are you comparing to a posterior spinal instrumented fusion where we have very low reoperation rates at 2 and 5 years, or are we comparing to some of these nontraditional growing surgeries where we've accepted historically very high rates of device failure given the challenges of the patient population?

So I would say the classical definitions of revision, removal, reoperation, and supplemental fixation is challenging in the pediatric world and oftentimes the investigators ourselves, we don't know whether this was disease progression or device failure or what category it falls under when we lose control of the growing spine and have to go back in and do more.

That being said, we've been really pleased to have some HDE approvals and ways to study devices prospectively despite not having a full PMA. And I think there are areas of the adult practice that probably fall under this category, as well, where there are higher reoperation rates, in order to achieve potentially better function and at the end of the day, how is the patient doing and how much did that impact the final outcome when the child had to go through two surgeries or 10 surgeries instead of one, are they a functional individual, are they able to walk and run and play and hold a job.

And at some level, for the complex cases, I think that final outcome maybe outweighs the reoperations that were entailed in the interim. But all this is very deserving of a study and as you said, you can't talk about complications without talking about outcomes. So with that, I'll pass it off to the rest of the group who can share the adult perspective.

DR. DEVLIN: This is Vincent Devlin, I just had one addition to the excellent comments by Dr. Larson. I would say subsequent surgical interventions would not

necessarily always be considered a failure because in some cases, in the context of the treatment of a particular condition, it would be a stage in the entire process of care of the patient, whereas in other disorders, spinal disorders, it would be of much more severe clinical impact. So I think that's an excellent point which Dr. Larson brought up.

It would be helpful if we could progress this slide to the next slide, if that would be possible. Thank you.

DR. LARSON: And again, finding that -- what did you intend to begin with, did you intend there to be more than one procedure or not? You could see how this could become a bad trend, though, if every single trial was set off saying well, this was supposed to be a multistage procedure and look what happened. So finding that right balance of what's going to be multistage and expected reoperation versus what is unplanned reoperation, it's a challenge.

DR. ANDERSON: And another issue that came up in one of the trials I was involved with was we had an intervention that was done under local anesthetic, so it was a very low surgical invasiveness. The comparative was a laminectomy and fusion. And so the reoperation rate of this minimally invasive procedure was a lot higher than the laminectomy and fusion. However, the complication rates on the initial surgery was far, far lower and so you were doing kind of a minimal operation and accepting a higher revision surgery rate.

And so I don't know if that was good or bad, I mean, people who didn't like the procedure said oh, look, so many people had to have revision surgery, but on the other hand, they didn't have any major complication putting the thing in, like cauda equina syndrome and surgical site infections and other things. So there is a balance between what you're trying to accomplish, particularly when you're changing devices, as Dr. Larson mentioned, it's been a huge revelation with the newer growth rods and tethers, but you may end up buying other complications that you have to now deal with, but you still are

maybe probably better off. It's hard to know that if you're trying to say everything is an apple and everything is an orange.

DR. LARSON: We wish in that original study you would have had a planned reoperation a hundred percent of the time and then look, only 5% of the people needed it, right? But that's kind of a backwards way of a surgeon to go about it, but again, since you're expecting a certain percentage of people to need tune-ups or mild revisions, can you build that into the study design from the get-go?

DR. DEVLIN: This is Vince Devlin again and another situation I have encountered, and would be interested in experts' opinions, is how do you distinguish when a reoperation is treating a consequence of the initial surgery versus the natural history of the disease which is being treated, so that can be very challenging.

DR. BRODKE: Yeah, that's a good question and it kind of parallels the comment that I was going to say, Dr. Devlin, in that when you go back and re-operate or revise, say an implant was put in and the case was to decompress nerves and put in this implant to stabilize the segment, whether it's diffused or not diffused. If the patient has some recurrent neurologic symptoms, it may be because the implant wasn't put in properly or wasn't working properly or there was a problem with it.

But it could equally or maybe even more likely be that the decompression part, which is not the implant part, just the other part of the same surgery, was inadequately done or there was progression, as you asked, progression may be of nerve compression following, you know, in the months or years following the surgery, which may be the -- especially the earlier revisions may be the fault of the surgery but not the -- and the surgeon but not the implant, per se. Conversely, it could be an implant-related problem, so trying to parse those two out is really difficult. It's much easier -- and likewise, adjacent segment degeneration, going back to re-operate on somebody at an adjacent segment, we

are trying to make a determination whether if that's related to the first surgery or the placement of implants or the fact that implants are even in place or whether that's progression of disease. And that's very hard. It's sometimes very easy to determine that, if mal-positioned implants are involved, but sometimes it's hard to determine.

DR. KEBAISH: The other thing I think maybe can flaw some of the studies is in adults, when we look at disc replacements, (a) it's -- I mean we're looking at say, a 3-year follow-up. I mean, we don't have to always revise patients, so that may skew the results where people don't have to get the revision within the 3 years. So I don't know if it's a perfect system to capture revision within the period of time required say, by the FDA to say the revision rate is low or high.

So it goes both ways and it doesn't always capture what you're trying to capture and I don't know if there's a way around that. Clearly, patients are lost to follow-up and end up -- may end up getting a procedure elsewhere, as well, but that may be easier to track because they're lost to follow-up, although we don't know the reasons. But when a condition doesn't necessarily require immediate revision, that can clearly affect the results, as well.

DR. ANDERSON: Yeah, I think if we've learned anything from the arthroplasty people is that the registry or real-world data was much better than the preclinical testing in simulators and the randomized controlled trials which did not show differences in reoperation rates, let's say, for metal-on-metal total hips at 5 years, but the real world, once they were released, did show differences.

So I think in answer to what you just stated, the real-world evidence from registries is maybe a better way to look at it, at least for this particular question about reoperations. And I agree with Darrel. Occasionally, one would say yeah, I'll blame that one on the procedure, but most of the time if it's adjacent-level disease, who knows if it would've

happened or not. We know it happens with non-operatively treated patients, with patients treated with foraminotomies, laminotomies, people treated with fusions, people treated with arthroplasty, they all get, to some degree, adjacent segment disease, so we don't know if that's natural history or related to anything.

DR. LARSON: I think a good take-home point, as we design these registries, is that it would be helpful to get a few data points from the surgeon as far as planned versus unplanned reoperation; device related, not device related. I've worked with some administrative datasets or also some of the study group registries and sometimes it's hard to tell what that reoperation or revision was. So again, if we have an EMR data capture, trying to give one or two data points the surgeon's going to contribute could be extremely valuable because again, the registries can get larger numbers of patients over longer periods of time but the reoperation and device-related complications is sometimes hard to tell from just lists of codes.

DR. BRODKE: And to Khal's point, the issues around the 2-year time frame are somewhat artificial, we see short-term complications perioperatively in the 90 days to 6 months-plus time frame and then we see some longer complications that may be treated by observation for a while with the spondylolisthesis trials. You know, Fishkin's paper sort of classically shows that nonunion wasn't the problem early, but if you followed them out long enough, many of them need a revision.

And so understanding how patients are doing at 6 and 12 months for the early complications is one probable piece. The 2-year doesn't really solve that, that's really probably longer to solve the longer -- to really understand the revision requirements for what might be failures and that's probably where registries have to come in, or long-term follow-ups, as we've seen in many of the IDE study follow-ups.

DR. ANDERSON: Yeah. The spine is unique compared to extremities in that our

reoperations can be at the index level or an adjacent level and this was one of the problems classifying them. Knowledge of both levels is of great interest to spine surgeons because I like to know how often am I going to re-operate at the same level, how often am I going to operate at the adjacent level.

But in an FDA trial we had the situation where a patient had a plate on a fusion, the fusion was solid, so they had to operate at the next level and they were going to do a fusion or even an arthroplasty, you had to take the plate out. So that got docked as both an adjacent-level and an index-level operation, which struck me as strange because you weren't really re-operating at the motion segment, you're just taking a plate off so you can operate the next level. So better definitions of those, what is index, what is adjacent level, even though it seems pretty obvious, becomes difficult when you're in trials.

DR. GOLISH: So this is Ray Golish. The point that was made by Darrel Brodke about the need for long-term follow-up is very well taken, it's normal to have PATH 1 look like 5-year follow-up of your 2-year data for the experimental and control cohorts and that's a very reasonable thing to want. But again, even though PATH studies are so valuable, they do become difficult for the sponsor, I won't use the loaded term "burdensome," but they do become difficult to get that valuable data depending on what hypothesis you're testing and then that goes double because you really want a real-world conditions-of-use study as another PATH. So the here's another plug for moving those studies closer to a registry paradigm in order to make them doable.

DR. DEVLIN: In the spirit of simplicity, do any of our experts have any suggestions on an alternate method of classification for reoperations other than the one shown on the screen, which is the current CDRH guidance?

DR. LARSON: I think something more along the lines of infection, device failure, disease progression, would be easiest for us to sort out because this almost feels like we're

slicing the pie so narrowly you could take 10 surgeons in a room and we might classify a reoperation differently under this algorithm.

DR. ANDERSON: The other problem you get into is what about non-operative interventions but they're still interventions, you know, pain stimulators, even injections, dorsal root rhizotomies. In these FDA trials you see a lot of those go on for people who are not having satisfactory outcomes. They see a pain doctor and all of a sudden they get a whole series of these things and what kind of attribution are those? Are those reoperations? That's another category of interventions, let's say, and they're not just medical interventions by medicine, but these are actually invasive procedures that have potentially life-threatening complications. And I don't know how we'd classify those.

DR. KEBAISH: Are we currently able to distinguish a planned revision from unplanned revision or unplanned staged procedures like, for example, if you're going to supplement your instrumentation with an additional anterior procedure or posterior procedure, is that still categorized within the revision, as well?

DR. DEVLIN: I think it would depend on the context of the specific study in which you're describing, whether it was pre-specified or not, but those types of things could be worked out through the pre-submission process in the planning phase for a study.

DR. ANDERSON: Yeah, one example might be percutaneous fixation for fractures where I am planning on taking the rods out. So when I look at my results, I have to report, like 95% of the time I take the hardware out so I have to have a revision surgery, but that was the plan going into it.

DR. LARSON: Or you can have the flip side where you do a non-fusion anterior spinal instrumentation or vertebral body tethering and then that is not giving enough correction, you need to come back and do a fusion, that could be under this definition, supplemental fixation fusion after a non-fusion procedure.

DR. KEBAISH: I mean, we currently -- even in our ethics system we use an additional tab when you're entering your operative report to include whether a planned surgery or revision procedure is going to be required, so I think that may be something that would clarify that if there is an additional piece of information that's required to be entered on all procedures.

DR. HARNER: I'd like to jump in now and say thank you to the panel for an outstanding discussion and now we'll move on to our second question and answer session. Dr. Devlin and Dr. Jean will now moderate our second Q&A session. They will be joined by Dr. Jianxiong Chu, Acting Director of the Division of Biostatistics in the Office of Clinical Evidence and Analysis. Dr. Chu oversees a group comprised of more than 70 statistical reviewers and has an extensive experience in clinical trial design, real-world evidence, biostatistical and epidemiological analyses to help advance medical device innovation and safety. Please remember to send your comments and questions to the OHT 6 e-mail.

Dr. Devlin.

DR. DEVLIN: We have one question for the panel relating to assessment of pain. The question relates to potential options for including pain in a registry and is there a way that you've seen medication use and medication changes at specific time points as recorded as a surrogate measure of change in pain?

DR. ANDERSON: I'll start since I brought up the issue of pain. I think one of the secondary outcomes of all trials ought to be opioid use. If you go back to lumbar disc replacements, one of the striking outcomes was, even in what was "successful treatment," 60% of the patients were still using opioids 2 years after surgery. I somehow think that's not a success. But that may have been the times back then where everybody was given opioids. Nowadays, I think including that as part of a trial would be useful to show efficacy and how you want to define opioid consumption I'll leave to the FDA, they should know

more about that, but it's obviously best with some kind of morphine daily equivalent kind of analysis, but I would strongly encourage that. Getting pain in registries is very problematic because no one really uses the same definition or same question for pain. And patients, even if you ask them to average their pain over the last week, they somehow don't seem to take that into account and I think you really need a pain inventory of sorts or probably better, and I'll let Darrel comment on the use of PROMIS tools for pain assessment would be much better than a single question. I don't believe a single question is accurate at all.

DR. BRODKE: Yeah, I agree with Paul.

DR. GOLISH: So Paul --

(Cross-talk.)

DR. BRODKE: I'll just go ahead and then stop quickly. I agree with Paul, I think pain is a challenge and there's -- it's very subjective, which is part of the challenge, and it's not visible, which is another part of the challenge, that is to the practitioner, but it is very real and we absolutely need to be paying attention to what happens with opioid intake, it is part of the national psyche right now.

We do need to understand what happens to patients in their opioid intake based also on what it was prior to that surgery because it greatly affects their ability to get off of drugs and the speed with which that happens if they are either on or not on opioids beforehand, and so there needs to be some sort of specific reckoning of where the patient enters the study in terms of pain and opioids if we're going to be measuring success post by amount of and whether they're still on opioids.

DR. GOLISH: Yeah, this is Ray Golish. I agree with those thoughtful comments of Paul and Darrel, and I'd layer on this thing which is, in CDER the use of special adaptive -- like analgesic products, the use of the classic pain metrics, the VAS and the NRS are standard and it's not hard to -- anywhere but those are normal tools, their psychometric

properties have been investigated extensively. Within CDRH the tendency is closer to functional outcomes, especially ODI, NDI, and some stenosis inventories. But even those are anchored in pain. When you look through their questions, of course, question number one for ODI is pain and then all the other questions feature the word pain in the answers. And so we're all trying to feel different parts of the same element and the psychometric properties of those have been studied extensively, as well. We'll get to talk more about that.

But I think that, you know, accounting for analgesics and opioids, I'll revisit as well, as the kind of big unanswered question. You could use it as a secondary outcome, you could use it as an inclusion/exclusion criterion, you can use it as a gating variable like SSIs, survivorship from SSIs, if you increase beyond a certain threshold, or you could try to control for it as a covariate. All those things have been experimented with, I don't think there's a consensus answer there, but we need to work on that.

DR. LARSON: From a pediatric perspective, we will use the Scoliosis Research Society questionnaire, we'll use pediatric PROMIS scores, so I think patient-reported outcomes are being collected fairly routinely in the pediatric world. One thing I note is that as you do a study, if you only have 30 or 40 patients, I think you really can't expect to see much of a difference, there's so much inter-person differential, inter-age. As teenagers mature, their SRS scores degrade as they become adults. Most 10-year-olds tends to be pretty happy. And so I think you need often four or five hundred patients in a study to really see differences in SRS scores in our literature and many of the IDE or post-approval studies are not probably powered to look at those outcomes.

DR. CHU: Hi, I enjoy you guys' very nice discussion. I have a quick question from a clinical perspective about pain score. Usually I see, sometimes we see sponsor using like a VAS score and sometimes call a numerical pain score, so from your point of view, which one

is more valid?

DR. ANDERSON: It depends on how you're collecting them. If you're collecting them on paper, either is pretty easy, you know, a patient just draws on a line and you measure how long that it is and that's the VAS; whereas if you're doing it electronically, a numeric rating score is a lot easier because you put a series of dots and they can put that there. I'm sure there are some minor psychometric differences, but I haven't found them to be of any real difference in how you measure outcomes; using either method seems to be valid.

DR. CHU: Um-hum.

DR. ANDERSON: It's really what the question of pain is, is the more important one. Is this pain you're having right then, is it the worse pain you had in 2 weeks.

DR. CHU: I see.

DR. ANDERSON: And I remember I had a patient who had zero Oswestry and he had seven on his pain score and I went and met him and he was really happy with his result, he says I feel great. I said why did you report seven? He says well, 2 weeks ago I was lifting something out of my trunk and I felt a twinge of pain, it was about a seven and it lasted for 3 hours, so he wrote that down.

DR. CHU: I see.

DR. ANDERSON: And that's why I hate a single question of VAS because I don't think it's reliable.

DR. CHU: Yeah, so I just share my statistical perspective on the VAS versus numerical pain. Relevant to Dr. Larson about sample size, because when you design a study from a single point of view you like to ensure adequate power to detect any significant difference, if there's such a difference between a group. So the advantage of VAS is kind of continuous variable treated as 0 to 100. A numerical pain is categorical. So oftentimes if you use a VAS score probably you gain more power, but here you have a smaller sample size that's

required, so that's only another reason I asked the question. Thank you.

DR. ANDERSON: But it is a lot harder to use VAS than a numerical rating score.

DR. LARSON: And you're getting more --

(Cross talk.)

DR. BRODKE: And there's also a bit about patients being confused on the VAS, it also takes a bit of understanding the patient to put a panel on a dot that's unmarked or a dotted line that's unmarked. And I agree with you, I think from a statistical standpoint that it does make a difference, VAS is really more powerful from a practical standpoint, especially if we're moving to electronic data collection, it's much harder. You know, we've tried to do electronic with this -- well, we have version of it. It's hard. So I don't think it gives really enough, has any more information, and I think that the issues that Paul brought up around the question you're asking are probably much more important than the style in the end.

DR. KEBAISH: Another thing with the prevalence of electronic methods of collecting data, including pain scores, I mean it would be more helpful if we're able to capture anatomic areas more accurately where patients are able to pinpoint to where that number is because, as was mentioned earlier, I mean it could be that their hip hurts because they have osteoarthritis when they had a lumbar procedure, so trying to figure that out and most scores or scales are not geared towards specificity of where the pain is.

DR. ANDERSON: Yeah, that's a very good point. We collect patient-reported outcomes now and the one adverse effect of that has been -- we've gotten rid of the pain diagram, which I used to love to look at because I immediately could tell "oh, that's a radiculopathy or pain's outside the body, I ain't going to operate on that patient," and now I don't get that anymore. I haven't seen a good electronic version. There are ways that physicians -- we can draw in where the patient's describing the pain but haven't been successful at having a patient use a mouse or something and draw a pain diagram, I'm sure

that's possible, but I haven't seen that in practice.

I don't know, Darrel, if you've had any experience with that?

DR. BRODKE: I haven't seen it routinely used. I would love to have that, I agree, I'm missing the time when a patient was drawing pain as the entire picture way outside and everywhere and you understand a lot more about that patient than just their pain.

DR. ANDERSON: Patients could do it with emojis now.

DR. KEBAISH: And even those, like, crazy diagrams, it was like multiple different -- like you said, it tells a lot about these patients, too, whether -- you know, negative or positive, but it's -- I think it was helpful.

DR. JEAN: Hi, this is Ron Jean. I'm going to jump in with another question that we have from our inbox to the spine surgeons. When you are reading the labeling for a new medical device approval, how important is it to see the clinical trial results, especially results related to adverse events and subsequent surgical interventions in such granular detail in the device labeling documents, and what role do these labeling documents play in your clinical decision making regarding use of a specific spinal technology?

DR. ANDERSON: Maybe I'll take that one on because I have taken a lot of the FDA SSED reports and then rewritten them to report them in the literature so that surgeons can understand them. So I think they need to be reanalyzed so that surgeons can understand what all that data means. Right now there's just too much information in those reports. It's not selective to what's of importance to our patients and to the surgeons when selecting a device. I think it's fine the way you report it, it's absolutely great because there's a lot of data for people like myself to analyze, but it has to be translated into usable form, I would say.

DR. BRODKE: Yeah, I would follow on and say that I prefer to read Paul's paper than the actual labeling because I understand it better. It's a little clear to me, it's written in a

form of a study that I've been involved with before, engaged in and understand it. Not just Paul's paper, obviously. I actually prefer to go to the actual publications around the subject that are published in the literature to really understand the depths of it. I appreciate that it's in, agree with Paul on all aspects.

DR. LARSON: We also like to see things presented at meetings and have kind of open academic discussion, whether virtual or in person, and I think for surgeons to know the investigators and have a chance to interact with them and then to see the published literature, that's kind of more of a complete package as far as wishing to adopt a new technology. And obviously, we want to see good results in the peer-reviewed literature.

DR. GOLISH: This is Ray Golish. You know, there's an art that FDA practices in collaboration with sponsors and also consultants like Hogan Lovells and MCRA, who've been on the Webex, which is writing the indications and intended use statements, and that's the portion of the label that is most likely to be surgeon-facing if they're going to be digesting that. And I think you guys do a nice job of balancing, for one thing, the inclusion/exclusion criteria of the trial, but if you wanted everything there, you just put the inclusion/exclusion criteria, but you're trying to capture the difference between efficacy and effectiveness in the intended use and indications of use population statements, but then also boiling down specific adverse event conclusions that result in contraindications, which is important, too.

So those key things of intent of inclusion/exclusion criteria but also major safety events that get boiled down into one paragraph, intended use, and indications for use statements are the key and then everything else can be kind of synopsized into additional portions of the label for the interested reader because, as I said, we like to educate ourselves clinically in different ways, meetings, peer-reviewed publications, and as already pointed out, the meetings are helpful, that's a different way to consume the information, synopses, as Paul was saying, different way to consume it. And then as Darrel Brodke was

pointing out, peer-reviewed publications, they give a different look. I personally like to read the FDA's SSEDs because they show what the Agency and panel were thinking at the time of that decision making. Of all the information that's available, it's huge, what were the key things that were driving that signal, and peer-reviewed publications will get that through a little bit of a different lens. So in summary, it's all helpful but the effort and the art that goes into an intended use and indications for use statement is valuable and is most likely to be surgeon-facing.

DR. JEAN: So no, that's some great feedback. I guess maybe a more pointed question, is there anything that we're missing, either in our labeling or our decision presentation?

DR. ANDERSON: I think the labeling is obviously written with a lot of lawyerly lingo which tends to turn off surgeons and as Ray just mentioned, there is an art form to make it really useful for the practitioner. I realize the labeling is there to lay out kind of a legal protection on what this is, but it also has to be useful to the readers, such as surgeons, and we may not understand the regulatory lingo or how to interpret certain ways things are put, so you have to keep that in mind. It's much like asking our patients questionnaires, we have to have a certain linguistic style at the eighth-grade level when we ask these questionnaires. Well, you can't use regulatory language in your documents and expect surgeons to read them and understand them and use them.

DR. JEAN: And again, we are just thrilled to have all of you assembled here for this workshop. I guess, are there more specific critiques you could provide to us related to either our labeling or summary documents that might enhance the experience for either the patient or the surgeon when reviewing those materials?

DR. ANDERSON: Yeah, I have a question about the data that informs the basis of the submission. For instance, if you have an NIH-funded study, you have to make that data

available in a repository for any other investigators, obviously de-identified and stuff. The same thing is not true for FDA, that's between you and the sponsor. Obviously, the sponsor could make it available like Medtronic did through the BMP/Yale YODA study. But how often do you see that? Is that part of your regulatory options, to tell the sponsors hey, we can put this data over here so other people can use it or maybe we can put it over there in 3 years to give you a 3-years' jump start and then other people can use it, because the studies that came subsequent to the YODA investigation were very important studies and the same thing would be true for like the SPORT trial. You know, the SPORT trial initially was three trials but now there's like 300 publications on using that data because they made it available.

DR. JEAN: Another good point, Dr. Anderson. So I think what needs to be kept in mind is sort of the purpose of those documents. And so when we look at a regulatory submission, we treat their submissions as proprietary information and confidential. However, what we attempt to do when we write our labeling or complete our summary documents, whether it's an SSED for a PMA or SSPBD for an HDE or even a decision summary memo for de novo, is really try to give a good snapshot of what the basis for a decision is within the regulatory criteria for that particular program.

And so we do sort of present aggregate results and do go into some details, when needed, whether it's a specific revision or what procedure was actually performed after the device failure. But in other categories, for example, the AEs that we've been talking about, we present that more at a sort of global study level. So it's not a question that we actively do in every case wherever we direct a company to figure out a way to disseminate that information in even further detail. With that said, there are business agreements that are made at times with different companies that are able to leverage datasets from a prior study and you'll see different examples of that. But it is definitely something that is

dependent upon if there's any public funding related to the study or if it's truly all on the part of the company, whether it's an OUS study, so there's a lot of considerations that go into that.

DR. GOLISH: So Ron, this is Ray. Let me maybe answer the question a different way, as opposed to individual elements that might be added to labeling is just the format of the information. People who study such things tell us that nobody reads things anymore. I'm getting old fashioned, I like reading SSEDs, but maybe I'm one of the few. So that information in there is so valuable and you pretty well distill it already. If it were made animatronic, flashy, listenable or watchable, I think people would actually enjoy consuming it in a way that gives them the scientific gist to key decision points that were driving the decision because the main points of safety and efficacy are pretty distilled. Maybe you can get the sponsors to do it somehow.

DR. KEBAISH: I think one thing that I would want to see, but don't know if that's possible, is postmarket data on those devices at some point in a follow-up because a lot of things are -- that we really don't see in the initial IDE and they show up later on, so how do we get that? I'm not sure what the logistics currently are, but it would be helpful to add those to the information whether it's labeling or just the website.

DR. ANDERSON: That's a great point, I totally agree. I do want to mirror Ray's comments that the SSED reports that the FDA puts together are superb and they're a huge resource, and I congratulate you because that can't be easy to distill all that information in somewhat of a digestible form; certainly for researchers it's very digestible and you are to be congratulated and I wouldn't change that too much.

DR. DEVLIN: This is Vincent Devlin. Continuing on the topic of labeling, we've had some additional questions come in. In addition to the reoperations and the discussion of adverse events, the question that came in relates to a new element that's been included

more recently, meaning patient preference studies and data. The question is how is this information perceived by clinicians on the panel and how is it utilized, and do they have any recommendations for the future in this regard?

DR. LARSON: This is an area of interest for me. So I have been working on shared decision making for some of the choices adolescents make and their parents make regarding their scoliosis treatment. It's been very eye opening to collect data from the parents' perspective and the patients' perspective, and I've learned that the families sometimes care more about how long they'll be out of sports and how much time in the hospital and what's going to be the immediate impact on their life. As a surgeon, I really am focused on complications and reoperation, but oftentimes the parents and families are looking at different things. So I think it's key, I think it's key to get information from your patients, what do they care about, what risks are they willing to accept, and I applaud the FDA for supporting this line of work.

DR. GOLISH: So this is Ray Golish. I think that it ought to be widely recognized, as FDA does, that this is critical, has long been and is getting the attention it's long deserved. It falls under the general umbrella of patient-centered outcomes and patient-centered studies. We know that PROs have worked hard to incorporate things that are patient centric into their so-called content validity, we'll get to talk more about that later, I think there's always room for improvement but it's worth noting and, I think, celebrating the Division of Orthopedic Devices, CDRH has really been in the vanguard of incorporating PROs as key points of their primary endpoints for a long time. Other agencies, other -- groups, other parts of the FDA have really learned from the OD and CDRH in incorporating PROs, so you've been in the vanguard with patient preference items, especially the very simple question, "Would you do it again?" the under-asked question but always valuable question, fits right in with the history of leadership around PROs.

DR. DEVLIN: Thank you.

DR. ANDERSON: Yeah, I have a comment. We created three questions that we ask every patient to write down when we see them, basically asking them to describe what you want out of this visit, and what was surprising is in more than half the cases it was not what I anticipated they wanted. Some of them just wanted a second opinion, some of them wanted medication, some of them wanted surgery in the afternoon, and it was very insightful when you ask the patients these questions. But I think it has relevancy when you're going to the clinical trials because if the patients, what they're requesting does not match with what the device or the trial is going to do, that patient may not be a very good candidate to be enrolled in that trial. So that should be taken into account, at least as far as the inclusion and exclusion criteria.

DR. HARNER: Any further comments?

(No response.)

DR. HARNER: With that, then, we'll end this Q&A session, an excellent session. Thank you to the panel. We will now go into our last break and we will resume back here at 2:20 p.m. for the final part of the workshop. Thank you.

(Off the record at 2:10 p.m.)

(On the record at 2:20 p.m.)

DR. HARNER: Welcome back to our final session to close the workshop. Our next clinical discussion will focus on the evaluation of effectiveness for spinal devices and will include case presentations. We welcome back Dr. Anderson, Brodke, Golish, Kebaish, and Larson. We are also joined by Dr. John Benson, a neuroradiologist at Mayo Clinic. Dr. Benson has a strong research focus with emphasis on vascular imaging, spine procedures, and temporal bone radiology. Dr. Moazzam, Dr. Devlin, and Dr. Sachs will be introducing the discussion topics.

Dr. Moazzam, please introduce the first topic.

DR. MOAZZAM: Hello, good afternoon and welcome back to the FDA. The first topic for this afternoon's discussion of effectiveness will be regarding study endpoints, specifically, how to select clinical study endpoints. Our questions that we pose are as follows:

What study endpoints are most meaningful for assessment of study success?

Number 2: Are regulators measuring what is most important and what matters most to patients?

Question 3: What patient-reported outcomes, clinician-reported outcomes or performance criteria, such as imaging outcomes, need to be assessed to define success in a spinal device clinical study?

And finally, Question Number 4: What should be the timing and duration for endpoint assessments?

DR. GOLISH: Hi. So this is Ray Golish, I've been asked to give some introductory remarks for this point of the session and I think I can be a little bit provocative in my approach to moderating here because a lot of rich information has come before this session and two great discussions are coming after this on radiographs and around PROs, so I'm going to be provocative by stating a firm opinion on the question that was asked and maybe put a target on my back and get to debate it with my colleagues here. I do like to play the devil's advocate, so I'm not going to claim that I'm completely wed to this opinion, but it is highly defensible in an evidence-based way.

And so the answer to the Agency's question about are regulators measuring what's most important, the answer is yes. Now, there's always room for improvement. We'll talk about some areas for improvement around the control of pain, identification of opioids, and then even more patient inclusiveness, as has already been discussed around patient-

centered outcome items. But there's a reason why the answer is yes, is If you look at all the IDE spinal trials that have occurred in the 21st century, the ones that are public information have either resulted in a PMA or been taken to panel regardless of whether it resulted in a PMA. You compare a couple dozen of them, you get essentially what emerges as a consensus design, a de facto consensus, not an official one. Now, of course, there's variability in the designs, but many look very similar regardless of whether it's a fusion design, an arthroplasty design, or motion preservation, flash stenosis or a reherniation design.

It's important to note that the consensus exists not because FDA is proscriptive. For people who don't interact with the Agency in this way, FDA does not tell a sponsor "thou shalt do this" and that's sufficient. It's negotiation about what constitutes safety and efficacy, valid scientific evidence, and is least burdensome. So when you look at all that and then look at the consensus design, it's very common to see a five-variate composite endpoint, a composite endpoint measured at 2-year follow-up with less than 90% lost to follow-up, very often in a non-inferiority paradigm with comparison to an active surgical control, so of course, not always.

It's important to note the five variables in the composite include both multiple safety and efficacy endpoints. Dr. Devlin had a nice slide early in the presentation which showed those essentially five endpoints, each of them dichotomous, meaning a binary variable, and then they're combined into a composite and that composite very importantly includes a logical "AND" patient-wise. For each individual patient to be considered a success, they have to be a success in all five variables. Now compare that to the so-called co-primary endpoint in which the patients are logically "AND-ed" but group-wise, not individually, you can also compare that to what's a more typical composite in which you take a logical "OR" of each patient's binary variables. It's really important to draw that out because in some of

areas of clinical science and biostatistics, this design, consensus design, is somewhat atypical, a single endpoint, co-primary endpoints, "OR-ed" endpoints are more common.

Within DoD and CDRH, it's been very common to have a five-variate composite endpoint, which is the logical "AND." There's some good reasons for that, because each of those things are getting to the question of what are the important categories and components is critical. So the five variables are PRO success, absence of secondary surgical intervention, the absence of neurologic deterioration, the absence of device or procedure related SAEs, and the radiographic outcome.

Now, when you look at the PROs, we're going to get to have a rich discussion with Dr. Brodke about this, but I want to be clear, in the lumbar spine, ODI has emerged as the PRO of choice. It's not just because it's the only one around or was used previously, there are some good reasons. Remember, the three main categories of psychometric properties for PROs include reliability. ODI showed strong-to-very-strong test, retest, and inter-rater reliability in multiple studies; not just psychometric, but clinical trials.

The second category, sensitivity. Keep in mind that the nice thing to note about ODI is not that it's been used in thousands of studies, although that's true. The nice thing to note is it's been used in multiple, major, very high-quality regulatory trials and in trial after trial you see a decreasing exponential curve showing an intervention decrease in the PRO that is shown to be sensitive for a known effective intervention.

And then, of course, the last property is validity, especially the controversial and complex concept of content validity. This is the most debatable portion of any PRO because it's an a priori spine criterion which says are we capturing everything the patient thinks are important, and a modern intellectual trend that's gotten very rigorous with focus groups. For any PRO that's 40 years old, not all that rigor will have been followed early on, but when you go back and look at those properties in terms of content validity, it has been done

RMDQ, you can see that they're reasonably strong and that there are no pathologies, so the other two, reliability and sensitivity, that is introduced by that.

And then finally, I'd say Dr. Brodke has sort of the best study of PROMIS for spinal interventions recently. RMDQ is a good alternative to ODI, but when we look at PROMIS, he showed nicely that there's a diminished floor effect and a diminished time to completion. But what I see in that beautiful research is that the diminished floor effect is mitigated by inclusion/exclusion criteria for randomized trials, but also PROMIS has an adaptive administration feature which is helpful in efficiency. The best thing about the PROMIS questions, really, is that they look very similar to the ODI questions.

Finally, when we look at radiographic criteria, I'll put out as a marker the CT scan at the final follow-up time point of 2 years is essential. Sponsors have done other things including plain radiographs, both in the spine and in the peripheral skeleton. That's really fusion or motion preservation bridging bone on all three columns; that attempted fusion should be considered the gold standard.

The absence of exercise is an important variable because it's a gating variable. Remember, the best analogy for this is looking at cancer patients. Imagine you have a trial, a cancer trial, the patient has an outcome for mortality and a PRO. If the patient actually expires, you can't measure the PRO totally, right? But in the spine, we could measure their PRO, we just ought not to want to, so we need to combine that variable in a survivorship statistical analysis right up front as a combination.

They even talked about the absence of neurologic deterioration, it's an important safety signal, and the absence of device and procedure related SAEs, the question here is whether to put those SAEs in the primary endpoint or not. Of course, they're studied always as a safety endpoint. So in the end, open questions that have already been discussed that are very important are how to deal with concomitant analgesic opioid use.

This is a very challenging question. The possibility of a secondary endpoint, of using it in the inclusion/exclusion criteria, using it as a survivorship variable like the SSI, using it as a covariate to control, those are all relevant discussion points. And then how to further the use of PROs in addition to ODI, not instead of ODI, to capture patient preference items with a number of different psychometric possibilities there. So those are my opening thoughts and I look forward to hearing from the Agency and my fellow panelists.

DR. BRODKE: Ray, I thought your comments were perfect. I agree with most of them. And I'm not sure that I'm going to -- I'm going to pick one spot that I might just disagree, only for discussion. Again, I'm not sure how strongly I disagree with that spot.

But before I get to that, let me get to the third question on this list has to do with PROs and clinician-reported outcomes and imaging outcomes, do they all need to be assessed, and I would argue that they actually measure different things and that we do still need radiographic outcomes and we have a whole section coming on that.

As an important sort of corollary outcome, because that's -- a lot of what we do is something -- is a physical success measured on radiograph separately from the patient reported success that we get in feedback from them, and they don't always correlate, we know that in spine, we've seen it for decades, but both are important. I don't want to lose one to the other. I absolutely think we need to be very thoughtful around the patient's perspective and we'll be getting into that a bit when we talk about PROs later, also, but I just wanted to underscore that at this point and then maybe push back a little bit on timing. The 2-year endpoint seems somewhat arbitrary to me, as we've talked about, and it's not quite long enough to identify some of the important failures that we'll see later and probably too long for most of the regular outcomes that we're looking at. There are very few changes that occur between 1-year and 2-year almost every study. And so I'm not sure I wouldn't argue to give up the 2-year time point for a subset of longer follow-up, so a

1-year standard plus longer follow-up standard, kind of a combination piece rather than the current 2-year standard. So I'd throw that out there just for discussion.

DR. GOLISH: And that's a great point, Dr. Brodke, I'm sure any sponsors out there listening are silently applauding at their desk. And I think maybe originally, historically, part of the energy of the time of 2 years was to capture that pail of fusions that aren't apparent at 1 year but are at 2. But one trial, in fact, cervical spine, had a primary endpoint at 1 year, but by the time they had all those data submitted, they had the 2-year data in any case, so your point is well taken and that is, I think, particularly true with past studies that routinely go out to 5 years, they're still able to look at that signal.

DR. ANDERSON: Ray, I'd like to comment about the neurologic deterioration as an endpoint. I think that's critical, that's what differentiates spine from other orthopedic things, is we have to worry much more about the neurologic dysfunction. The problem with the way the definition is now is a patient may have subjective numbness in their hands that they didn't have pre-op and now that is, first of all, an adverse event and it may also mean, as an endpoint, it may fail that criteria. So our ability to utilize some of the neurologic exams is really beyond how are accuracies to measure it, as well as the meaningfulness of it.

Another example is reflexes. I think reflexes are useless, they're very subjective. Somehow people think that they're not, but they're totally subjective. I never agree with a reflex exam between me and my resident, and I certainly never agreed with Darrel's neurologic exam, and so I think that's a useless criterion. And I don't even know what's good or bad. Is a reflex disappearing good? Is a reflex getting worse bad? I kind of think of a four-plus reflex as a bad thing on most spine things. So you don't even know how to interpret them, so I'd get rid of the reflexes. The sensory is just too subjective. I think if you really wanted to do it, it would have to be specified with a pin, some kind of pinpoint

sensation or two-point discrimination, but that's just incredibly time consuming and I don't think it's all that valuable, particularly given the plethora, especially in the cervical spine, of people having all their nerve neuropathies, median nerve neuropathies, who knows what else, comes in complaining of numbness. It's so non-specific to the spine, it's not very helpful. More exams can be a little bit more objective, but again, that goes to the quality of the examiner. Are you going to have trained nurses do that or are you going to have orthopedic surgeons and neurosurgeons do that or residents? It's problematic.

I think gross categories of neurologic function such as the Frankel scale or modified ASIA are very important, for sure. Maybe the ASIA motor index, which is a bit more subjective with 10 key muscle groups graded zero through five, could be used. But if you just have a four biceps instead of a five because your shoulder's got a rotator cuff disease or something, is that meaningful or not? So I think, again, that neurologic criteria is just too narrow because it doesn't account for the noise that goes on in people's lives. That definitely could be revised.

DR. LARSON: And again, bringing it back to clinical practice, so in pediatrics it's not clinical practice to do a pinpoint needle exam on every child at every postoperative scoliosis visit. So again, having the study parameters match what's currently being done in clinical practice will make the parameters more meaningful and then also make the study more feasible to complete.

DR. BRODKE: Also likewise, patients will frequently report success and happiness with their procedure yet still have radicular numbness, just not radicular pain or weakness. And so I think, to Paul's point, pain and maybe a motor assessment are probably the most important two pieces to that, but that certainly the rest of it is subject to some concern, both for the variability as well as the distance from what the patient cares most about.

DR. ANDERSON: And you probably need to do an attribution. If somebody has a

neurologic deficit, is it possibly related to the cervical spine or is it related to something else? For instance, it was like 10% of the cervical disc arthroplasties had lumbar spine surgery, which I presume they would occasionally have neurologic dysfunction from that. And so that needs to have an attribution and I don't think it currently, at least the trials I was reviewing, ever did that, but that would be something to consider.

DR. LARSON: I'm interested in patient-reported activity levels, as well. I mean, we treat some people who are level 7 gymnasts and win their state competition and we do surgery on them, do they go back to doing that level of gymnastics? And similarly for someone else at a different point in life who could drive a car and make it around the grocery store and sleep through the night without pain. That's really the role of the patient-reported outcomes, and having good patient-reported outcomes as far as activity assessment, quality of life, are they at the same functional level that they were preoperatively?

DR. HARNER: Thank you.

DR. GOLISH: Thank you, Dr. Larson. I can take that and extend it a little bit to say what about the concept of more objective functional tests? More objective, unlike our neurologic exams; functional, accounting for baseline, as you said, in stenosis and neurogenic claudication, walking tests to the merge to stay somewhat more objective though functional, and patient-centered measures. So what are the panelists' thoughts on that going -- embracing the patient centeredness of PROs, but making them a little bit more objective and reproducible?

DR. HARNER: In the interest of time, I think we're going to have to move on to our next topic. Now we'll turn things over to Dr. Devlin to begin case presentations. Thank you for the excellent discussion.

DR. DEVLIN: Thank you, Dr. Harner. Good afternoon, everyone.

The next session deals with the assessment and definition of spinal fusion in spinal device regulatory. In this session, the effects of spinal devices on fusion assessment will be considered through a series of case presentations. After the presentations are concluded, this will be followed by a discussion about the important factors which impact the assessment of spinal fusion in spinal device clinical studies. We'll present five cases and each case will be presented by the clinician who provided us with the case. The first case is presented by Dr. John Benson. Please advance the slide.

DR. BENSON: And one more time, please. So yeah, first of all, thanks for having me.

So these are the preoperative images. So you can see, first of all, that the area of interest there is at C4/5 and we spotted a lot of changes and then quite a bit of spondylolisthesis. And so we also get flexion-extension dynamic images that are showing translational motion on flexion-extension, so there's increased spondylolisthesis of C4 and C5. Sir, can you advance again, please?

So now we have the post-op images, and I think you can take the left and center images together because they're both done at 1 year out. The radiograph does show interbody ankylosis, although it's harder to see on that than it is on the CT, the CT that clearly shows it. But I think one thing, one reason I wanted to show this is that all the ankylosis you're seeing there is within the cage itself and there's nothing dorsal to it at all. So you know, by some criteria, this might fit as an adequate fusion and by other people's, it might kind of raise some concern. But in any case, at two and a half years after the surgery, we have a pseudarthrosis that certainly declared itself at that level.

And that's all I had to do to introduce them.

DR. DEVLIN: All right. Next, we'd like to go around to our experts on the panel and have everyone provide their observations and comments on the assessments and the presentation.

DR. BRODKE: Maybe I'll start. I'm looking at the 1-year X-ray and I'm looking at the up, which is upright, and a saggital CT scan which is supine and they look different to me. I got a 1-year upright X-ray, there's a slight listhesis that I don't really see on the saggital recon of the CT. The bone graft in the PEEK cage is spotty looking on X-ray, the ossification, and where it looks almost confluent and solid on the CT. And then you move to 18 months later and then there's a clear cleft through the cages and level of the disc and lucency around the proximal screws, and you know for sure that it's a pseudarthrosis. I don't see this very often, what I think becomes a clear fusion that later is recognized as a pseudarthrosis, but it looks in this case like this is the situation in this particular patient.

DR. DEVLIN: Thank you.

Dr. Golish.

DR. GOLISH: Dr. Devlin, I'm having trouble seeing the images on the presentation, so I will pull them up on my computer and get back to you.

DR. DEVLIN: Sure. I think if you click on the small picture or double click on it, it may show up in your screen.

DR. GOLISH: Yes, okay. I have that. I did get a chance to see this case earlier and I agree with everything Darrel Brodke just said, you know, to feel completely certain of that, given what he pointed out. Is it an atypical clinical course as time evolved? I would need to see the full set of images, meaning the entire computed tomography. But as an atypical looking cage suggests, what he said is a late nonunion. If that's the case, we'd want to investigate a reason why that atypical evolution -- and at the top of my list would be some very low-grade indolent infection that I'd want to know about, whether it be by biopsy, inflammatory markers or some other additional imaging.

DR. DEVLIN: Thank you.

Dr. Anderson.

DR. ANDERSON: Yeah, I think there's two attributes of fusion we should consider. One is bone bridging across the inner space or if it's a posterior fusion, between the lamina or across the facet joint. And secondly, is there motion on dynamic X-rays? Dynamic X-rays could be flexion-extension but it also could be, as Darrel suggested, difference between supine and upright.

In the cervical spine, this topic has been discussed ad nauseum at the Cervical Spine Research Society. Probably at the time Darrel was president of that society, the society did an evidence-based review to look at what is a fusion, and they found that with extremely high sensitivity and excellent specificity, that if there's less than 1 mm of displacement between the spinous processes on a flexion-extension, then that patient clinically and radiographically had a fusion success.

So I think that kind of criterion in the cervical spine is very useful and I think it's very evidenced based and I don't think for regulatory studies that we need to do CT scans. Now, obviously, if clinically you suspect a nonunion, that might be the way to go to confirm it, but for a regulatory thing, I think I'd be very happy with what the CSRS came up with as far as their definition. I don't know if you have any comments about that.

DR. BRODKE: Yeah. And clinically, a lot of people use that, as well, and it's a bit of a continuous variable in terms of its alignment with fusion and that it's very clear, the more motion you have, the more likely there's a nonunion and less motion, you know, zero to one is likely a fusion, 2 mm or more is now pretty much guaranteed a nonunion, and it's the equivocal sort of in between those. And I think many of us use that clinically for follow-up rather than getting CT scans. CT scans do help, but they can fool you, as well, as they did in this picture. In fact, a well-packed fusion cage, a post-op day one CT scan can look fused because it looks like bone can be traversing from one vertebra to the next, if really tightly done. So there are some challenges on CT scans, even though we think of them as the gold

standard.

DR. DEVLIN: Thank you.

Dr. Kebaish.

DR. KEBAISH: Great points, obviously, about validating the fusion and I think we all have been fooled, maybe more myself, looking at the radiographs and even CTs. But at the end of the day, I think the clinical picture needs to be correlated with the imaging and you can see here, this patient had a CT at 1 year. So clearly, the patient had some symptoms that warranted further investigation. We don't do CT in someone who's asymptomatic.

So that really maybe adding some additional correlation with clinical presentation to the studies and maybe that warrants additional studies as opposed to just -- I can tell you, relying -- you know, in cervical spine may be more so, we're relying on the flexion-extension in someone who's having back pain, it's probably going to fool you seven out of ten times or maybe more. So I think clinical presentation is very helpful and hearing you towards doing more or not. But you could see -- I guess it's not predictive, the CT at 1 year, but there clearly is some reason for having that CT, it was done at the time.

DR. DEVLIN: Thank you.

Any comments, Dr. Larson?

DR. LARSON: I think it just highlights the complexity of that 2-year follow-up mark of being the gold standard that is multifaceted and sometimes things look pretty good at 1 and 2 years.

DR. DEVLIN: All right, great.

DR. BRODKE: I also would underscore what Khal said about marrying the radiographic and clinical pictures because really, one without the other doesn't help us and having criteria that calling it a success with one without calling it a success with the other may fool us.

DR. ANDERSON: The other comment I would have is a lot of submissions now are on biologic devices where what they're looking to claim is they're increasing the speed of fusion and since even at long-term follow-up we have a hard time describing the fusion, I have no idea how you tell that somebody fused at 6 weeks instead of 8 weeks or 12 weeks, and personally, I reject those kind of arguments out of hand because I can't measure it at any time point.

DR. BRODKE: If it would be to sacrifice the patient and look biologically.

DR. ANDERSON: Yeah.

DR. DEVLIN: All right, great comments. Why don't we move on to the next case and the next slide, please.

DR. BENSON: Sure. And then please advance one more time. And I think this is a good example to tie up on that last conversation because now we will see the motion on it. So this is a 3-year postoperative exam; it's a two-level, obviously. The top level, to my eye, looks good. It has good interbody fusion. And then the bottom level, I can see a horizontal lucency extending through it. And then advance one slide, please.

So now we have flexion-extension. And now going back to what was just talked about with the interspinous distance, we can see that happening here. So at the fused segment at C5/6, the spinous processes really essentially do not move at all on the flexion-extension, but you can see them splaying on the flexion view. And I've seen different things used for the criteria, I've seen zero or less than 1, which essentially is the same to me. But if you get much more than that -- I can't remember who said it, if it was Paul -- but if you get a 2, then you're kind of looking at a nonunion there and again, we're seeing that cleft here, the horizontal cleft in the same spot. So advance one more time, please.

And this is just kind of the confirmation, so now we do have a CT and again, it's just confirming what we were expecting on the radiographs where we were having that clear

interbody ankylosis on the top level, but the bottom level has a clear pseudarthrosis there.

DR. DEVLIN: Great. Well, thank you very much, John.

Why don't we just go around quickly and see if there any additional comments.

Dr. Larson.

DR. LARSON: I think it would be interesting if this patient had no symptoms. Is that still considered a failure? And then again, how do we tie in that patient reported outcome piece? Or on the contrary, if this person has a longstanding narcotics history and has been on disability for the last 10 years, is that a nonunion and a failure? That's all I have to say.

DR. DEVLIN: Thank you.

Dr. Kebaish.

DR. KEBAISH: Again, just looking at this, and I think there are different challenges with each device, like I don't know if we're showing other ones later on, but there are some devices with those incorporated fixation or degraded fixations, which some of them may be challenging depending on the interbody portion of it. You really, again, have to figure out whether you want to keep digging into it, and I've had to go and get bone scans on patients with long fusion just because I was suspicious that an area had not fused when the CT scan was not clear or did not give me a clear answer. So it still is the same comment I made earlier with clinical, and PROs suggest there's something not right then maybe further studies should be done.

DR. DEVLIN: Thank you.

DR. KEBAISH: Merging studies, I mean.

DR. DEVLIN: Okay, thank you.

Dr. Anderson, any additional thoughts?

DR. ANDERSON: No, I think this is an example, an excellent example showing the usefulness of the splaying of the spinous processes and how useful that can be. There's no

doubt, I think two-thirds of pseudos in the cervical spine are relatively asymptomatic, so it's not unusual to be asymptomatic with this. But this would fail radiographic criteria of a fusion in the trial, but it may not fail the clinical outcome, it may still be satisfactory.

DR. DEVLIN: Great. Does anyone else have any different comments that have not yet been expressed?

DR. GOLISH: So this is Ray. I'll just comment that again, getting to the points about correlation of radiographic and clinical outcomes, that can't be overemphasized and Dr. Larson just mentioned that.

In the consensus design, for a patient to be a success, they need to have all variables met patient-wise, so they can fail radiographically or clinically in terms of an NDI PRO or both. We all know that we have cervical nonunion patients out there who feel like a million bucks, so that's wonderful, clinically. But in order to get the best assessment of those, you know, for the purposes of really a pivotal IDE trial, the criteria are so tight because we really do need to be assured of a plausible causal mechanism, meaning union, which underpins our treatment paradigm, or a plausible causal mechanism, meaning radiographic motion preservation for an arthroplasty that underpins our clinical treatment paradigm and our PROs. So I think we need both for pivotal trials and I think that the radiographic look has to be comprehensive, CT and flexion-extension, just at that final primary endpoint.

DR. DEVLIN: Great, thank you so much.

Any other comments before we move on to the next slide?

(No response.)

DR. DEVLIN: Why don't we go on to the next slide, please. Okay, now Dr. Anderson will present his case. The next slide.

DR. ANDERSON: Well, so this is a patient who, for a spondylolisthesis, underwent anterior lumbar interbody fusion with a titanium cage and then percutaneous posterior

instrumentation, and this is what it looked like at 3 months. The patient was doing well, we were high-fiving each other and everything looked really good. If you look at that implant, you see a really nice border between the bone and the implant, this has got some surface coating on it, so we're hopeful that bone is in-growing into that thing. You don't see a lot of bone through it at this point. Let's go to the next slide.

And here's a CT scan at 3 months and again, one of the problems with any cage is there's usually some kind of mismatch between the endplates and the surface of the cage. You could see that that cage was a little rounded on the top, it might've fit a little bit better. At the bottom, it really looks like bone is juxtaposed to that endplate and to me, this is looking really good. On a coronal view, it actually looks like bone is remodeling through it, but as I think Darrel said, boy, you pack that cage full and put it in and you have a great carpentry, it looks healed the day you put it in. But to me, is this thing healed or not? Well, it might be. Three months seems awfully early for an interbody fusion, however. Let's go to the next slide.

And here we are, just more saggital views. Again, to me, everything looks really good and you kind of get the idea that there's bone forming within that cage. All right, let's go the next slide.

And 2 years, sure enough, this patient did extremely well. Now you can see probably bone behind the cage, it looks like there's bone within the cage, there's really no screw loosening, so this looks healed radiographically. And the point of presenting this is, you know, what do we need to assess healing? Particularly when we have metal cages now, is a CT convinced, do we need CT scans or do something like this on plain radiographs, if I graded that, and we'll let Dr. Benson review it since he's a radiologist, but I can't make that look any better.

Any comments, Dr. Benson?

DR. BENSON: Well, a couple. First of all, this is a great example, I think, of how well sometimes you can see this on radiograph. But compared to the cervical spine, I think it can be more challenging, particularly in large patients, to get this good a finish. So this is a great example that you're showing, but it's not always this satisfactory in terms of looking at the interbody ankylosis on this.

I mean, we can never forget that the other things that we're seeing here, especially on these last images, is really reassuring stuff, and the things that we're not seeing, right, we're not seeing any sort of vacuum cleft, we're not seeing any sort of hardware failure or subsidence or loosening or anything like that, so there's a lot of other criteria that you could use that you would still potentially be satisfied with even if you did have a more challenging patient that was harder to see the ankylosis itself.

DR. ANDERSON: Yeah, I'd like to point out to people, the problem with the coronal view is that at L4/5 and L5/S1, those discs are angled substantially different and during surgery, when we're implanting this, we rotate our C-arm or use navigation to account for that so that we're avoiding parallax where the X-ray beam is parallel to the disc space and if we're doing L5/S1, that's quite a lot of angulation.

But when we go to assess plain X-rays 3, 6 months, the X-ray tech positions it and aims it somewhere. Ours aims it at like L3, which is not even close to the angles that we really need and that really limits our ability to do any kind of measurements or even assessments, because if you look at this cage, you can't tell what's going on in the cage on that AP X-ray because the X-ray beam is not parallel to the disc space. And I just threw that in just to point out the challenges of plain X-rays, especially in the lumbar spine where you have this significant lordosis going on.

DR. BRODKE: I would add to both of the comments, all the comments, because they're really quite right, and I would add that when we see cross-trabeculation like we do

outside the cage, like in this case, on the lateral film we see it posterior to the cage, bone growth from one endplate to the next, we can be much more comfortable calling this fusion on an X-ray but we don't always see that, to Dr. Benson's point, and I think that's probably maybe because some implants allow this more than others and maybe some surgical techniques allow this more than others. But I think when we can see either anterior or posterior to the implant crossing fusion mass, it really helps.

And the lack of lucency around the screws in sort of a corticated line between the lucency and the bone, so you often see the vertebra and then we see this corticated line, then lucency, then the screw, which is a clear sign of a pseudarthrosis and we saw that on the last patient, two patients. We don't see that here. It's 2 years later, you would start to see that.

So I think those are really important pieces and we'll often see -- when we see bone anterior or posterior to the cage or lack of lucency, we're much more comfortable saying this is fused than the alternative, as Dr. Benson said.

DR. ANDERSON: You also don't see lucencies around the cage to the bone, that's really important, too. I would point out that the seen bone behind the cage could be the iliac crests are there or it could be osteophytes. You can be easily faked out because it looks like bone right behind the cage; you think it is, but it's really not. It's due to overlying shadows. That's again part of the imaging problems at L4/5, is you're shooting through the pelvis and so it's adding another difficulty in your assessments.

DR. DEVLIN: All very rich --

(Cross-talk.)

DR. GOLISH: Paul, this is a great case. It looks so nice. My only critique is that I never high-five a lumbar fusion patient at 3 months, but hopefully you got the high five at 2-year final follow-up here.

DR. ANDERSON: Yeah, at least they haven't started a malpractice suit yet.

DR. DEVLIN: These are all great comments. Unfortunately, to leave us time for a discussion, it would be important that we move on to the next case. The next case will be presented by Dr. Kebaish. Could you please advance the slide?

DR. KEBAISH: So this lady is a 46-year-old female who underwent the surgery elsewhere and she came to see me 5 years, actually 6 years to be exact, after this procedure. She stated that she had some pain relief early on but since then she has had some back pain and she was basically scheduled to have a spinal cord stimulator. Now, you look at the CB X-rays here and you see the cage, the integrated fixation with what looks like some graft anterior. Now, be mindful that this has been more than 5 years since her surgery. And do you have the CT? We had some -- can you show the next slide?

These are her radiographs, these are the CTs, and you look at the CT and I mean, they look pretty good. I mean, the screws actually look solid and I don't see lucency on the screws. You can see some bone across both in the coronal and saggital plane. We did flexion-extension, I don't see them there, but there was no motion really noted on the flexion, but she was a poor flexion-extension patient basically because she was having pain. I mean, maybe we can discuss here, what does the panel think about this CT and radiographs and about the evidence of fusion or status of the fusion?

DR. BENSON: Well, I'll jump in first. You know, contrary to what I just said last time, at first blush I would say when you look at things that are not present, everything kind of looks reassuring again, right, you're not seeing lucency, you're not seeing any sort of hardware failures, subsidence, or anything like that. And obviously, on the CT it looks like there's evidence of interbody ankylosis there. The one thing I would bring up, and I still see this sign debated in the literature when it comes up, is at 4/5 there's what we call the sentinel sign and that's usually a radiographic sign, but right now we're seeing the CT

correlate for it, and it used be that was evidence of solid fusion across the site. But I think more and more now what we're seeing, people argue that that's just a sign of hypertrophic bone changes, a real bone spur that actually might be a sign of non-fusion. And you can see that it is anterior, there is that little lucency between that, right, it looks like this little chronic defect that it's not actually fused against it. But if you're going to make that argument, you might say that this is at least one sign that it's not fused at the L4/5.

DR. KEBAISH: How about L5/S1?

DR. BENSON: At L5/S1, I don't see anything like that. You know, I can't tell on the coronal if you're trying to show me a little gap between the hardware and the endplate or not. If you're not trying to show that, if that's just kind of an odd bone averaging, then I would say that that looks pretty good. And again, it's hard to just rest on that one sign. I don't know if it was there preoperatively. And if everything else looks good, it would be hard to just hammer this home as any sort of nonunion seen there at L4/5.

DR. KEBAISH: Dr. Brodke.

DR. BRODKE: Yeah, I agree with Dr. Benson's assessment. The anterior sentinel sign, if you wanted to call it that, isn't really -- at L4/5 isn't confluence, so I don't think that qualifies indicating fusion, but cross-trabeculation and boning reaction on each side of the implants at 4/5 and 5/1 look quite good on both planes. I have a hard time calling this a nonunion from the CT. There may be other ways to do this, dynamic films or something. I know that she's not able to do it, but sometimes sitting or supine and then standing is enough to look for a change. But at any rate, I can't see anything that makes me think she's got a pseudo based on what I'm looking at here.

DR. KEBAISH: Dr. Anderson.

DR. ANDERSON: Well, as far as I could tell, I agree with Dr. Benson. This looks like it probably healed. But again, you need to look at the whole CT sequence, both -- I

particularly like the coronal and the sagittals and just scroll back and forth to see how big is that bridging bone that we see in the center of the cage because at L4/5, the only thing that's healed is what's going through that center of the cage and oftentimes that's kind of like a narrow, small pillar and maybe it's insufficient. It's surrounded by metal, which maybe is going to give you some volume-averaging artifact. I'm not exactly sure what that word means, but radiologists use it a lot so I thought I'd co-opt it, but as an excuse for them to be wrong is kind of what I interpret that as. If she was symptomatic, I'd be really concerned that this hasn't healed, but I don't see anything that's pushing me to do anything.

DR. KEBAISH: So I mean, I wasn't trying to be deceptive, but actually I submitted the full CT scan on video as well as intraoperative images and this just really was trying to highlight the difficulty in assessing these patients. This patient had very gross motion, more so than you'd expect in someone who hasn't had a fusion and actually, her intraoperative images were very telling. And again, that again stresses the point of clinical picture and that really pushes us even to go further than the imaging themselves, where the CT didn't look particularly convincing for a nonunion but her intraoperative images were a lot more telling and her symptoms were, as well.

DR. BENSON: And I'm trying to think if I've ever seen a study where this would not pass the criteria set up in the study, to act as a fusion.

DR. KEBAISH: Yeah.

DR. BENSON: You know, I've never seen anyone throw out anything for a sentinel sign. So yeah, I think this a great example of clinical correlation, for sure.

DR. BRODKE: This shows the weakness of any one radiographic image in defining failure.

DR. ANDERSON: And again, I go back to fusion means bridging bone and lack of

motion. So unfortunately, we couldn't detect lack of motion because of her inability to get those X-rays. I'm wondering, on the coronal, is part of this cage PEEK material, because there's lucencies or notches on the anterior endplate of L4 and the anterior endplate of L5 and I don't know if that's plastic there or is that actually --

(Cross-talk.)

DR. KEBAISH: Go to the last slide. I think it's metal, I think it's all metal. Go back to the last slide.

DR. ANDERSON: That would concern me if you see that lucency.

DR. KEBAISH: Um-hum.

DR. ANDERSON: That would concern me that there's motion there or a nonunion there or, as Ray suggested, I suspect many of these nonunions are cult *C. acnes* kind of infections.

DR. BRODKE: Yeah, I don't know if this cage per se, but it looks to me like there's a PEEK component to it or some radiolucent component to it between the front and the back columns on the lateral X-ray and then there's this central metal spine, if you will, to the cage which may be what's making the CT look like it does through this so-called volume averaging you brought up, Paul.

DR. BENSON: And then going back to Paul's earlier points, this is another great example of, first, on the lateral of the iliac crest you could kind of confuse the picture and second, on the AP, how you essentially cannot see any useful information at L5 through S1.

DR. KEBAISH: I think another factor that I -- a case I'm suspecting is the fact that those integrated fixations do move, so they don't loosen. So probably the motion, even though there would've been otherwise motion that would loosen the fixation, it's not showing up here because they just move at the interface with the device or the interbody device.

DR. ANDERSON: One of the concepts I've been interested in is stability over time and that is the implant stays in the same position relative to the bone over a period of time and defining that as healing, similar to a porous coated hip implant where once it grows in, it doesn't subside anymore. And both of these cages subsided and I was wondering if you looked at serial X-rays, if you could detect subsidence. There are ways to analytically determine that for research purposes. They're not very useful clinically, however. But that would be another concept, is are the implants stable to the bone, because if they are, that kind of implies stability of the construct, at least in my mind.

DR. DEVLIN: Those are all great points. In the interest of our limited time, I think we should move to the next case, so if we could please advance the slides. And one more, please.

DR. BENSON: Great. So we're at 1-year post-op here and obviously, a multilevel posterior lumbar fusion and a couple things I wanted to point out. First of all, if you just look at the AP view you can see the dorsolateral fusion. Not as well as a CT, obviously, but you certainly get a sense of it. But you also get a sense that unless there's a huge lucency sitting there, maybe it would be a little bit harder to detect some lucency through it, but even what you can pick up is that even though there's no true hardware failure, there certainly looks like there's loosening at the top at the L2 screws. We're not seeing any sort of vacuum phenomena within the L2/3 disc, but it's still concerning that there's loosening up there. So now you can go to the CT.

So now I wanted to again show this case for a couple reasons. One, you get a sense on that AP view, the left-hand picture, just how clear the CTs can make a really dense dorsolateral fusion look, and it can pick up the defects in it pretty well. So now we've confirmed the loosening around the screws and now we are seeing this one level of pseudarthrosis at the top at L2/3.

DR. DEVLIN: I think we may have to cut our discussion short on this case so that we have time to discuss the topic as a whole. If you could please advance to the next slide. What we'd like to do now is have a discussion regarding what we can agree or disagree on regarding spinal fusion assessment.

What imaging modality or a combination of imaging modalities are most appropriate for spinal fusion assessment?

What is the optimal timing?

What imaging criteria are most important? Should they be quantitative or qualitative or both?

Are there any regional differences between the cervicothoracic and lumbar region?

And how about special circumstances and challenges such as prior surgery, imaging artifacts due to metallic devices or radiopaque bone graft materials?

Next, what is the role of clinical outcome data when assessing fusion?

And then lastly, what are some important differences between fusion assessment and clinical trials versus assessment in clinical practice?

So because we only have limited time, why don't we just go down the line across our experts and feel free to make any comments on any of the questions you wish to address.

Dr. Brodke.

DR. BRODKE: Thank you. I'll hit a few and then stop. I think we can often identify fusion, particularly if the patient is doing well, with radiographs, that would be the standing AP and lateral -- films or circle films and then dynamic films with the caveat that if there is obvious cross-trabeculation and no motion on the dynamic films and a patient's pain is improved, I would be comfortable calling that a fusion. And if you can't tell if there's inadequate motion or crossing bony lines or something that's unclear, then we'd probably move to a CT scan. From a time standpoint, I think the earliest I would probably assess that

is at 6 months and then moving forward to 1 year and further years out. And that is, I think, true of particularly cervical and lumbar, a little harder to use those criteria in thoracic spine where it's both hard to see motion and it's hard to see the fusion mass.

DR. DEVLIN: Thank you.

DR. BRODKE: So I end up looking for lucencies around instrumentation, which takes longer.

DR. DEVLIN: Great, thank you very much, Darrel.

Dr. Anderson.

DR. ANDERSON: Yeah, I think in the cervical spine, as we've discussed, we could use flexion-extension films. The bright time point for that, I don't know, I would guess 12 months would be reasonable, it could be 9 months, but 12 months would be a reasonable time point.

The lumbar spine, I think you have to use a combination of AP and lateral radiographs, they have to be centered appropriately, and then a CT scan if it's an FDA trial where you're evaluating fusion as an outcome. It's not going to be a perfect result. You probably will not be binary; you'll probably be "yes, no, and who the hell knows" kind of an outcome on that one.

I think setting up a checklist of what is a fusion, like Dr. Benson mentioned, you know, the lack of loosening, the interfaces between the hardware and the bone, subsidence, clefts, things like that, having a checklist and all of those have to be negative and then the positive ones, bridging bone and stability is to your best ability to measure, would be all criteria that can give you an outcome. Again, it may not be binary outcome, it may have three levels of outcome.

Now, the big question is should we include fusion as part of our outcome criteria? And I'm mainly talking lumbar, because we can't measure it very accurately, in my opinion,

as we've just discussed, and I almost wonder if we should give up measuring fusion as an outcome even though these fusion trials often, because we can't measure it and it doesn't make any sense to have an outcome, a trial with an outcome measure that we can't measure accurately or reliably.

DR. DEVLIN: Thank you very much, Dr. Anderson.

Dr. Kebaish.

DR. KEBAISH: Well, those are all great points and maybe I'll address the last point here, whether some -- you know, how do we assess fusion in clinical trials and clinical practice? I guess if our ultimate goal is to come up with devices that achieve what we're trying to achieve clinically, then it should be similar and as has been mentioned earlier, the clinical presentation of the patient should clearly be the driver of additional imaging. So I would say if a device is really being tested for its ability to achieve a solid arthrodesis and especially now I'm hearing about the timing of that arthrodesis, then I think we need to be certain that that is going to be achieved and that we should use both clinical data as well as imaging as we normally would use in a clinical setting.

DR. DEVLIN: Thank you very much, Dr. Kebaish.

Dr. Benson.

DR. BENSON: Well, it's a really contentious topic, depending on who you ask. If you just wrap those two questions up as one, even if you look at the American College of Radiology, the ACR, which puts out imaging recommendations for essentially every disease state, this is really conspicuously absent both in terms of what you should use to image it and how often you should image it after an operation has been done. And even here at Mayo Clinic, when I reached out to the spine surgeons, even in this smaller group, they had at one point tried to come up with a very regimented way of however one was going to do it and even there they couldn't agree with what they were going to do. So it certainly

seems to be individually based on how people feel and probably what kind of cases they've run across in the past. I think if you -- you know, you could come across a reasonable way to do it. I think Dr. Brodke had mentioned something like 6 and 12 months for starting out with radiographs and I would agree with that, certainly with flexion-extension, I think, as part of it. And then I think you could either use CT for ambiguous cases or you could just decide that, for a new device, you'd always have to get a CT at, for instance, something like 12 months. You know, CT is more accurate but it obviously comes at a cost, both a financial cost and a cost of additional X-rays. Radiation, too.

DR. DEVLIN: Thank you.

Dr. Golish.

DR. GOLISH: Yeah, I'll go quickly. Number one is I've advocated for CT, if not the last word, but it's the first word given all the complexities we have discussed. Optimal timing, I think 2 years remains the gold standard for lumbar spine. I think the real question is, and as Dr. Brodke suggested, that be accelerated to 1 year as the primary endpoint knowing that the correlation between 1 and 2 years are high, not only in the cervical spine but in the lumbar spine, as well. And actually, FDA clinical trials have the most wealth of data on that very question.

But what imaging criteria are appropriate, I'll synopsize that as bridging bone in the absence of lucencies for any column that was attempted to be fused. There are IDE trials that have been one column in the posterior lateral gutter, two in the interbody space and three in a 360 need for fusion in all those attempted fuse levels.

What is the role of clinical outcome data when assessing fusion? I'll say we've already discussed the patient-wise logical "AND" that you need. For any patient to be a success, you need the radiographic success and clinical success, and I love and embrace and understand the arguments about one versus the other but especially, let's consider the role

of arthroplasty. If the patient's a clinical success, you need to know that they're a clinical success because they have some motion, not because it's ankylosed, because otherwise you could just stick a bottle cap in there, right, your entire treatment paradigm is undermined, and we need to remind ourselves we're not just practicing clinical medicine nor even just doing clinical science, we're doing preclinical science simultaneously when assessing things radiographically, especially the condition of the device and the potential for wear.

And then the final thing I'd add is for motion preservation you need flexion-extension films for any arthroplasty device or any motion-preserving stenosis or a reherniation device. And as much as I feel that's true, the real challenge is what to do with that data. You really want to bake your noodle and do what some sponsors have done and look at flexion-extension films and the two-level arthroplasty, lumbar arthroplasty, and try to figure out what that means. I don't advocate using those data or cutoffs around them as a failure criterion because we don't know how they contribute.

Do you sum them? Do you have to have a certain amount of motion at every level? Does it have to compare to preoperative motion, meaning there are some people with lumbar devices who have hypermobility relative to their preoperative status, is that good? Is motion like money, more is always better? None of those questions have been answered and that's particularly true that FDA IDE trials have the most wealth of information, we just don't know what that means. So those are my answers to the questions.

DR. DEVLIN: Thank you so much.

Dr. Larson, the last comments.

DR. LARSON: Last, but not least. Certainly, the pediatric world is a little bit different and we rely primarily on plain radiographs. I love Dr. Golish's concept of the motion-sparing surgeries need a motion assessment, so we have done that in our non-fusion practice, obtained standard of care radiographs to assess motion at 1 year following a non-fusion

scoliosis procedure. One thing we just haven't hit on too much is who's doing the assessment and in our pediatric world, it's been interesting that some of the IDEs, it's the surgeon doing the measurements for a surgeon-sponsored IDE. Others are calling upon a team of radiologists. I think the industry studies have a contract research organization that's doing the specialized image measurements, but those people don't have access necessarily to the full clinical story. So somehow having a protocol or a perspective on who's doing the assessment and how are those assessments combined to deem it successful or not a successful result. But great discussion. Thank you, everyone.

DR. DEVLIN: Thank you so much. Just in closing, my last comment would be, in looking at the totality of the data that we look atfor FDA, a challenge we see arising in the future with real-world evidence is that postoperative imaging is discouraged for many reasons in the orthopedic field because of its lack of impact on clinical outcomes. So I think that's a topic for future discussion. But thank you so much, everyone, for your excellent comments. I wish we had hours instead of minutes to discuss this, but we'll have to call this session to an end and turn it over to my colleague.

DR. HARNER: Thank you. Excellent cases and discussion.

Dr. Sachs, please introduce our last topic for this session. Thank you.

DR. SACHS: Thank you, Dr. Harner.

Yes, so now as we've been attuned to a nice segue and nexus for patient-reported outcomes, as we're seeing them and considering them in 2021 and beyond, the questions that we pose and ask the panelists to address are:

What patient-reported outcomes should be used now for the assessment of various spinal disorders?

And the second question is what patient reported outcomes need to be developed for spine or the spine community for use as spinal devices studies go forward in the future?

DR. BRODKE: I'll open the discussion with a quick discussion of my own here. As we're running a little bit late, I want to make sure we have the opportunity to really discuss this, so I'll try to keep it as brief as I can. We've already been talking quite a bit about the value of the patient's perspective in understanding outcomes and while we frequently measure the things that we can measure easily, such as reoperation rates, readmission rates, mortality, etc., really understanding the patient's outcomes in a meaningful way are the future. We've been doing this in spine and really in orthopedics, as was stated earlier, for a long time, so this is not new for us but underscoring this.

The challenges are many and I'll talk about a patient a little bit later here to underscore that challenge. But patients present at different points on almost any scale that you could use and they end at different points on almost any scale that you can use, and using firm cutoffs is meaningful for some and not meaningful for others. There are many things that affect patient-reported outcomes, not the least of which was brought up earlier and that is comorbid musculoskeletal conditions can cause changes in the outcome scores. Other comorbidities, psychiatric illness is an example of that, can cause patient reported changes in patient reported outcome measures.

And there are a lot of problems with how to administer them. The ideal patient-reported outcome measure is going to be a minimal burden to the patient or the office and be easily reproducible, not repetitive, and be precise. We've kind of gone through this list before, but I'm going to underscore the value of burden or avoiding burden because it changes the outcome scores themselves, actually, until finding patient reported outcome measures that don't cause undue burden in doing these, not just for the cost of the study, but for the value of the outcomes that we're measuring. We need to pick outcome measures that provide full coverage with minimal floor and ceiling effects and that are understandable and statistically valid and obviously, understood by everyone.

So with that as a basis, there are three general types of patient reported outcome measures. The first is one that we're most familiar with and we've talked about already, is disease specific measures in spine. The most common ones are ODI and NDI, Oswestry Disability Index and Neck Disability Index, and they're in multiple languages and useful. The ODI was built for assessing function in patients with chronic back pain. It has subsequently been used by many of us for all kinds of back procedures with varying success because it wasn't really built for that but nonetheless, that's become sort of the standard.

There are many other outcome measures that you'll see both through IDE trials and others, such as the Zurich Claudication Questionnaire for stenosis, SRS-22 for deformity, which has something in it on how a patient feels about their body, which others -- and their posture and body image, which others don't include and that's important for certain types of spinal surgeries. JOA or MJOA for myelopathies, another example.

These are all disease specific measures built around the diseases in the spine and built with classical test theory from psychometricians, meaning each individual question is scored. It's not valuable unless you have all or almost all of the questions answered and have some mechanism for understanding the addition of the scores of each of the questions.

Another class of outcome measures is called general outcome measures. SF-36 is the classic, we've seen it in many, many studies, it crosses disease patterns which is great and very helpful for that. It's hugely burdensome, it takes a long time to fill out and isn't ideally used by itself, it can't really be used very well by itself, which will kill you in the world of spine care. There's shortened versions like the SF-12 which give us almost the same amount of information, but certainly not as accurate in any individual patient, but in large groups it can be. It again crosses disease patterns and so it's helpful for understanding how diseases work compared to other diseases once treated.

There's a different class called quality measures, like the EQ-5D or a modification to SF-36 called the SF-60, which really allow us to understand quality of life and understand changes in quality of life years, and I think is very valuable for understanding how we treat patients.

And within the PROMIS group, and PROMIS is a little bit different in a number of ways, but actually in this way it's not. PROMIS has a quality measure called the global health 10 or the global health 29, 29 or 10 questions depending on which one you use, and it can act as a quality measure.

And then lastly, we have domain-specific measures, and the best example of that now, as we know, is sort of the PROMIS domain measures like physical function or pain interference or depression or anxiety. Those are all domains of health within the World Health Organization and they've been split out by the folks that made the PROMIS outcome measures and the value of the way the PROMIS measures, all of them, were made is that they're all made in the same way so it's the same scoring system. There's a t-score, which is kind of the mean of the entire population, which is set at 50, and then one standard deviation of the population is plus 10 or minus 10 on the scale and every one of the PROMIS measures works the same way so you know what the score means as soon as you see it.

The value of ODI, as Ray said, is that it's the widest-used outcome measure, well understood by spine surgeons. It's pretty easy to administer, actually, and it's only 10 questions, so it's pretty functional. There are some issues around floor effect, but those are made up for by some of the other things I just mentioned. I think it doesn't give us a complete picture, personally, and I think we need to add -- I think we're in great value by adding a physical function and/or pain interference score from the PROMIS group. The real value in the PROMIS measures is they can be used in the computer-adaptive testing process so you can administer four questions and get the specificity of the entire question bank of

over a hundred questions, depending on which bank you're using. And so the burden is much lower than the addition of multiple other outcome measures. And then lastly, we talked already about it, so I won't spend any time on it, is VAS and numerical pain rating scales which provide yet a third mechanism for understanding outcomes. PROMIS, as Paul mentioned, has three versions of measuring pain, but the most commonly used is called pain interference which is actually not so different from the ODI except that it's built with psychometrics to allow for CAT, computer-adaptive testing.

With that, I think I would recommend -- just to throw it out there and let others respond to -- I would recommend including a disease specific measure like ODI, still because it's the most widely used and understandable. At some point you may be able to get rid of it, there's crosswalk tables now built around the physical function scale of PROMIS and ODI, so we don't absolutely need ODI, but it does provide to some people understanding of something that others don't yet understand, so it's hard to get rid of yet. The physical function scale of all of the PROMIS measures is probably the most appropriate for understanding how patients are really doing with their function.

And I was going to tell you, there's one anecdote, real quick, and that is a patient of mine who was a university gymnast, who at her worst had a physical function score of 50 -- 50 is the mean for the country -- and she was miserable and couldn't be a gymnast, but she scored 50 on her PF CAT. This is a problem with using raw number cutoffs. At any rate, she underwent surgery, her pseudos completely resolved and she went back to competing at collegiate level with gymnastics and her follow-up score was 72, over 2.2 standard deviations of where she was when she was injured and it -- you know, this sort of an off-the-charts good was her normal and average was her disease. So we have to be really cognitive of how PR is working in that regard. At this point I think I'll stop and let others comment. Thank you.

DR. DEVLIN: Thank you very much, Dr. Brodke.

If we could shift over to Dr. Larson, if you wouldn't mind, Noelle, to give us a brief perspective from the pediatric point of view.

DR. LARSON: Again, with pediatrics, some of our children are nonverbal and their developmental status varies over time. Asking a 10-year-old to fill out an Oswestry Disability Index is going to have a different result than an 18-year-old. The Scoliosis Research Society score has been used for thousands and thousands of studies, but it does have weaknesses including a ceiling effect and a focus on developmentally normal children. Often, that questionnaire is not applicable for children with disabilities who may not care how they look in clothes or be able to respond to how often they go out with their friends.

For the patients with developmental differences or under age 10, we're using a disease specific measure called the EOSQ, early onset questionnaire, which is pretty generalizable and asks about things like breathing and activity level and works for children who may be in a wheelchair. So the patient reported outcome has to be tailored to the device that's being studied.

I think one big problem again is that ceiling effect, that many of our patients are functioning at a very high level, as we heard from this gymnast or varsity level high school athletes, and we see a ceiling effect where patients can easily do the things that are listed. But I think we do change people's function with our spine interventions. After fusion surgery I have patients frequently comment they have more trouble swimming, they have more trouble checking their blind spot, but none of this is really being captured in the patient reported outcomes that we currently have. So I do think we need a better activity assessment for some of our pediatric spine patients, something like a Tegner knee score or something that again is going to capture what level of activity they are doing or perhaps, as Dr. Golish was talking about earlier, a score or a functional measure like a 6-minute walk or

a 50-yard dash, something that's going to capture how they're functioning in the environment.

DR. DEVLIN: Thank you. Thank you, great comments.

Can we move to Dr. Golish next?

DR. GOLISH: So the two presentations have been so masterful, I'd just like to ring a couple of bells that were embedded in them. One is Darrel's Brodke's point about the utility of computer-adaptive testing. It's hard not to like this, right? It is burdensome to patients to fill out questionnaires. If you can use the known statistic psychometrical properties to have them fill out fewer questions, we should do that.

I think one challenge with PROMIS is that technology is so appealing it gets conflated with the PROMIS measures themselves and then that gets conflated with -- as Darrel very clearly pointed out, the difference between a disease specific outcome versus a domain specific outcome, and although domain specific outcomes are appealing, I think one challenge is establishing whether the minimum clinically important difference is the same between the domains that exist in different disease states. I would need to see lots of measurements to demonstrate that that's true.

You know, estimating an MCED is hard, it's controversial, but it's absolutely critical to everything we're doing as we dichotomize every single PRO against MCED in order to get a binary variable to bake into the primary endpoint. And if your MCED calculation is at least not consistent between disease states, you introduce a potential new source of variability. So when I think about this, I don't think are the ODI psychometrics perfect, I think which metric has better psychometrics, and for the disease state we're discussing, the answer is none and RMDQ may be comparable. Now, as evidence accumulates for domain measures, I would just say whether the MCED calculations and all of the baseline characterizations of the domains are the same between disease states, that's a big open question.

DR. DEVLIN: Thank you very much, Dr. Golish.

We're really running out of time, unfortunately. I don't know if you have just any closing comments, Dr. Kebaish.

DR. KEBAISH: No, I think this has been a great discussion and I agree with all that's been said about the outcome. Clearly, most of what we do is to improve quality of life, so we need measures to accurately depict that. And the other thing is they need to be fairly easy to administer, so we don't want to go through the SF-36 on every patient where it's hard to collect a lot of data and really don't know what the utility of it is. I like the EQ-5D, we clearly need a clean measure. I think computer-generated ones are probably going to be more popular and I think they make a lot more sense. But again, talking about disease specific or condition specific outcome, we need to improve those and have a repository of outcome measures that have been tested and maybe in the future improved upon.

DR. DEVLIN: Great, thank you so much.

Unfortunately we're out of time. I'm getting the prompt, so thank you and we'll move on to the next session with Dr. Harner.

DR. HARNER: Great. Well, thank you. That concludes our last clinical discussion. We will finish the workshop session with a question and answer session and it will be 10 minutes long. It will be moderated by Dr. Devlin, Jean, and Chu.

DR. DEVLIN: Dr. Jean, would you like to take the first question?

DR. JEAN: Sure, I'll jump right in. We have a question: In the future, real-world evidence obtained from registries has great potential. However, in the near term, registry data has been most immediately applicable to support device labeling expansions, modifications, and postmarket surveillance. Can you suggest any additional scenarios related to spinal devices where real-world evidence would have utility in the near term, and do you have any thoughts on overcoming some of the median and long-term hurdles to

expand the utility of real-world evidence for other regulatory applications?

DR. LARSON: I can take this one. I think, again, in pediatrics we have small rare disease populations, so we discussed earlier that there's a real need for studying what the current care is. Unfortunately, the current landscape does involve some off-label or physician directed use utilities, I mean, so we would love to capture this in our registries and to be able to study what is being done out there in the big world. I know there's major regulatory hurdles for that but, from a surgeon perspective, we feel like that would make the world safer.

In the interim, certainly continuing with premarket and postmarket studies using registries is, I think, one of the peds ortho's visions because we do have strong registry presence and trying to expand those registries, making the barrier to entry lower for sites and keeping a high level of quality for the data, the registries hopefully will keep a high level of data that can be used for device monitoring and research.

DR. DEVLIN: Okay.

DR. ANDERSON: I have one comment, that perhaps utilizing the existing registries as a CRO to conduct clinical trials, for instance, contracting with the AAOS spine registry to conduct a clinical trial and that way that would maybe lower the cost for the sponsors, but it would also give the AAOS a lot of experience in collecting data that could ultimately improve and make it a bit more standard of care.

DR. DEVLIN: Great. Well, thank you very much, great comments. We have one more question remaining. The question is directed towards our panelists. Do you see a need for development of a new patient reported outcome measure to enable measurement of pediatric patients through and beyond the time they enter adulthood? If you agree there is a need for development of such a patient reported outcome measure, do you have any suggestions regarding specific elements to include in this new assessment tool?

DR. LARSON: We are hopeful that PROMIS will do this for us. The PROMIS metrics for younger children and adults are not that far different and our institution where I work has implemented universal PROMIS scores, so every time an orthopedic patient walks in the door, they fill out the PROMIS. Again, that remains to be seen. Certainly, the Scoliosis Research Society score is validated more for adults than for children, but it's been probably used more widely in children. And finding the right outcome score for every scenario probably is not feasible but for sure, there's nothing that physiologically changes about a child when it goes from -- when they go from being 17 to 19 or 20.

So studying that continuum and making sure what we do to children in childhood is paying off in their adult years is critical. So a great question, and I think there's work under way, for sure, to marry the EOSQ to the SRS score and then beyond that, I'm hoping PROMIS will serve this role for us.

DR. DEVLIN: Great, thank you.

DR. ANDERSON: I would like to say the same problem exists from an adult to a geriatric patient, is that the outcome like the Oswestry Disability Index I don't think works very well for a geriatric patient with spinal stenosis, who has intermittent pain or let's say, with osteoporotic vertebral fracture, I don't think those are the best outcomes because they can't answer a lot of the questions or are just not appropriate for them. So in the same way we need to have some age-specific geriatric outcome measures because, again, a geriatric patient is not just an older adult, but they have a whole different set of problems just like pediatric patients have a whole set of problems. And I know you are accepting the Zurich spinal stenosis scoring system, which I think is very good for at least that condition, but there are other conditions that we operate on. Adult deformity in a geriatric patient would be another example.

DR. BRODKE: If we're really talking about the future, PROs are going to be piece of

it, but there may be another way that we actually end up measuring the intermittency that Paul just talked about or how patients really are doing, because we don't always know just because of how they fill out their PROs or whether they're accurate or affected by other disease and that is activity monitoring. We may see activity monitoring become a future outcome measure, actually, that surpasses much of what we have today. There's a lot of different centers working on how to do this and what it means and how it correlates with both PROs and some of the other physician-related outcome measures that we've discussed or that we've passed today. But think about how that might really look in the future as a way to understand patient results is through monitoring their various activities.

DR. GOLISH: Yeah. Amen, Dr. Brodke. Dr. Larson and I were having a text-box conversation as to more direction and I like your phrase, "lots of places are working on this." It will take time to validate, but once it's validated, it will be a huge win. Efficiency and accuracy is something that is more objective but yet patient-centered simultaneously.

DR. LARSON: And making it fun for the patients, right? I mean, if we have it, it's going to be an app, it's going to be something where they get personal feedback that "I walked around the lake tonight." It's going to be patient-centered and enjoyable.

DR. DEVLIN: I think these are all great comments that outline the future. Unfortunately, we are out of time, so thanks to everyone who participated and I'll turn things over to Ronald Jean now.

DR. HARNER: I'll make a few closing remarks and then turn it over to Dr. Jean.

DR. DEVLIN: Please.

DR. HARNER: This does conclude our last session on the spinal device clinical review public workshop and again, I want to thank everyone for their participation. It has been my sincere pleasure to serve as your Master of Ceremonies for today's workshop. On behalf of the FDA, I would like to thank all of our speakers for their great presentations and scholarly

dialogue. To all of you in the audience, thank you for staying engaged and contributing to this workshop.

I will now ask Dr. Jean to provide the closing remarks.

DR. JEAN: Sure, good afternoon again. We've all had some incredible discussions at today's workshop and I know that we'll all be looking carefully at the proceedings again to see how we can improve spinal device clinical trials as well as taking this feedback into account within our divisions for these practices. Based upon the comments, spinal fusion is a bit more complex than one would believe at first glance, but the feedback provided gives us context to think about when developing criteria for assessment.

Similarly, we've had some great presentations and discussion on how we can overcome the challenges related to enhancing diversity in clinical trials. Admittedly, we've only had 1 day to have dialogue about these topics and we've only scratched the surface. I look forward to future work and make further progress on these issues.

I also want to point out that trying new things like having frank discussions as we did today or working in areas where we can improve or leading the way on PROs goes beyond the Division of Spinal Devices and is really an effort by OHT 6 Office of Orthopedic Devices.

There are a number of individuals to thank for the success of today's workshop. I'd like to first thank Captain Raquel Peat, my office director, who has not only provided direction and support, but is really visionary in leading the way in transformational activities like our OHT 6 series of workshops that are added value to the American public. Dr. Vincent Devlin, our office chief medical officer, an orthopedic spine surgeon, has been instrumental in providing expert guidance on the content of this workshop to ensure that we achieve maximum value. Our newest office clinical deputy director, Dr. Christopher Harner, was a wonderful Master of Ceremonies for today's event, and I appreciate his willingness to assist. I would like to thank our other FDA presenters, including Dr. Sydney Gibson, who presented

on behalf of Dr. Elizabeth Adegboyega-Panox, Dr. Hongying Jiang, and Mr. Colin O'Neill for their excellent presentations, as well as Dr. Dirk Alander, Dr. Caroline Moazzam, and Dr. Bart Sachs for introducing some engaging topics for our clinical discussions.

We would not have had a complete perspective on the topics discussed today without our participants from outside of the FDA. These individuals have not only added to the great discussions, but also have given up their time from very busy schedules to contribute to this workshop. I would like to thank Dr. Paul Anderson, Dr. John Benson, Dr. Darrel Brodke, Dr. Raymond Golish, Dr. Khaled Kebaish, Dr. Noelle Larson, Dr. Hassan Serhan, Ms. Janice Hogan, and Mr. Justin Eggleton for providing their unique perspectives as expert clinicians, scientists, and medical device industry consultants.

Last but not least, I would like to thank all of the OHT 6 regulatory health project managers, Ms. Sahlee Sabala, Lieutenant Commander Randoshia Miller, and Lieutenant Commander Ogochukwu Ogoegbunam for their incredible project planning behind the scenes, as well as Ms. Barbara Richards, Mr. Trent Knight, and the rest of the crew at the FDA TV Studios running this production.

I will close by thanking the hundreds of members of the audience gathered today who have given us your careful attention and posed some great questions. We look forward to hearing any additional feedback that you have for our forthcoming participant survey.

Thanks, everyone, and I hope that you have a wonderful afternoon and weekend. Stay well. Bye-bye.

(Whereupon, at 3:57 p.m., the meeting was adjourned.)

C E R T I F I C A T E

This is to certify that the attached proceedings in the matter of:

VIRTUAL PUBLIC WORKSHOP - SPINAL DEVICE CLINICAL REVIEW

September 17, 2021

Via Zoom Videoconference

were held as herein appears, and that this is the original transcription thereof for the files of the Food and Drug Administration, Center for Devices and Radiological Health, Medical Devices Advisory Committee.

TOM BOWMAN

Official Reporter

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947