ACTTION IMMPACT XXIV - Pragmatic and Comparative Effectiveness Clinical Trials of Pain Treatments

---

October 22, 2020

A Matter of Record
(301) 890-4188

ACTTION

INITIATIVE ON METHODS, MEASUREMENT, AND PAIN ASSESSMENT IN CLINICAL TRIALS
IMMPACT-XXIV

Research Design Considerations for Pragmatic and Comparative Effectiveness Clinical Trials of Pain Treatments

Virtual Meeting

Thursday, October 22, 2020
12:00 p.m. to 4:30 p.m.

C O N T E N T S

C O N T E N T S (continued)

P R O C E E D I N G S

(12:00 p.m.)

Introductions and Meeting Objectives

DR. TURK: Hello. I'm Dennis Turk. I'm the associate director of ACTTION. Bob Dworkin is the director of ACTTION, and we are placing this virtual program on for you as one of the ACTTION initiatives, and specifically the IMMPACT meeting. Before I formally welcome you, I'd like to go over a few details related to housekeeping.

This is the first virtual meeting that we've been doing, so therefore we're trying to make sure we accomplish things as efficiently as we can, and you'll have an opportunity to give feedback on that at the end of the program.

For general housekeeping, questions can be submitted throughout the meeting by clicking on the "Ask a Question" button located in the engagement panel. After each presentation, five minutes have been allocated to any questions requesting clarification of information presented.

All other questions should be addressed

Page 5

during the panel discussion. So therefore, if you have some questions about lack of clarity of something that was presented, then in fact you should raise that question at the end of the presentation or else save more substantive questions for discussion during the panel discussion.

Each panel discussion will include two presenters as well as two moderators. To participate in the consensus discussion, click on the "Consensus Discussion" button in the engagement panel. You will be directed to a new meeting page.

Per the updated publication policy of ACTTION, anyone desiring to be a co-author of the manuscript developed, based on the meeting proceedings, must attend both days of the meeting for a total of at least 6 hours. Should you require another copy of the updated policy, contact Valorie Thompson at the website that's on this screen.

A list of IMMPACT meeting participants and the 2-day agenda can be found by clicking on either

Page 6

the "Meeting Participants" or the "Agenda" buttons within the engagement panel. Please complete the post-meeting evaluation form that can be found on the "Feedback Forms" button within the engagement panel.

This is a really important point. We really want to learn how to do this effectively and efficiently, how to have these virtual meetings, so we will make use out of the information you provide when we start planning for the next virtual meeting that will be coming up.

Now, for the agenda that you're going to be covering, it starts at 12:00 Eastern Time and different times depending on what time zone you're in. You'll get an introduction from me with the objectives, which is what you're getting right now, then there will be some discussion for the first two presentations, discussing definitions and general considerations for pragmatic and comparative effectiveness clinical trials, and then some statistical considerations that one has to have in mind as they think about these types of

Page 7

studies.

That will be followed up by two presentations that will be examples of circumstances, tried and developed, to use pragmatic trials or comparative effectiveness trials, and then we'll have a break. After the break, there will be a systematic review of the methodological aspects of pragmatic and comparative effectiveness clinical trials of pain treatments.

Now, after those presentations, there will be panel discussions. As I said earlier, at the end of any one of those presentations, if you need clarifying questions, you can ask those at the end of the presentation using the chat button, but if you have more substantive questions, please save them for the panel discussion.

On Friday, we're going to begin, again, the same time, and there will be discussions of issues related to how you conduct these types of studies; types of outcomes; the eligibility criteria; using different sites; rescue treatments; et cetera, in the same way we talked about previously. Any

Page 8

clarifying questions can come after those talks.

After that, there will be a break and then a panel discussion. Importantly, following that panel discussion, there will be a consensus discussion where specific recommendations can be made for pragmatic comparative effectiveness trials of pain treatments; specifically, what can we recommend to someone who's actually planning to do one of these studies?

We mentioned IMMPACT. Many of you have been involved with IMMPACT before, but just to clarify for folks that are new to this, IMMPACT stands for the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials, I-M-M-P-A-C-T. It's an international consortium of participants from academic research, governmental agencies, and you see those listed there, industry, consulting and research organizations, and consumer advocates.

The mission of IMMPACT is to suggest methods for improving the design, execution, and implementation of clinical trials of treatments for pain. IMMPACT does not emphasize outcomes of those

Page 9

actual studies as far as trying to determine one treatment is better or not as well as some other treatment, but rather how would you design to improve upon the kinds of studies that we're doing to get the best benefit of those studies?

IMMPACT is part of ACTTION, which stands for Analgesic, Anesthetic, and Addiction Clinical Trials, Translations, Innovations, Opportunities, and Networks, or ACTTION. You'll notice we use lots of acronyms. The director of ACTTION is an expert on developing acronyms, so you have to bear with us as we use these.

But what is ACTTION? ACTTION is a public-private partnership with the United States Food and Drug Administration. The mission of ACTTION is to identify, prioritize, sponsor, coordinate, and promote innovative activities with a special interest in optimizing clinical trials that will expedite the discovery and development of improved analgesic, anesthetic, addiction, and peripheral neuropathy treatments for the benefits of the public health.

Page 10

Who is IMMPACT? There have been 23 prior IMMPACT names. This is the 24th. There have been over 200 participants at these meetings. Some have been at more than one of those meetings. Academic and related participants from multiple countries, representing over 125 different academic institutions and health systems, have been participants.

By participants, I mean they're engaging in the meeting, so not just an audience but they actually participate in the meeting; and investigators and reviewers from national and international government regulatory research agencies, including the FDA, NIH, EMA, MHRA, and many others. Those individuals actually participate in the study and in the discussions that are going to be conducted, so they're not just passive observers, but they actually are participants.

We have representatives from over 50 different pharmaceutical and device companies and consumer advocacy representatives from seven

Page 11

different organizations, and I haven't listed them all. There have been several individuals from different private consulting organizations.

But most important is you. You are IMMPACT, the people who attend these meetings. They're the ones who actually contribute to the discussion, the consensus, and the development of recommendations and considerations for future research.

What do IMMPACT and ACTTION do? Well, we hold consensus meetings in research initiatives. We publish those types of systematic statements and reviews. We commission actual papers on topics that we feel need additional attention. We conduct and support scientific studies. We sponsor the development of diagnostic classifications. We support educational initiatives.

There have been almost 150 IMMPACT and ACTTION articles that have been published; over 600 different scientific journals, ranging from addiction medicine to veterinary medicine and women's health since the inception of IMMPACT in 2003, and it's been cited over 12,000 different

Page 12

times according to Google Scholar. So we have meetings. We develop consensus recommendations or considerations. We develop manuscripts and publish those papers. They have had a reasonable impact. They've been broadly looked at from a whole range of different journals, and they've been cited fairly widely.

If you have more interest in knowing about ACTTION or IMMPACT, you can go to the website. You notice IMMPACT is I-M-M-P-A-C-T, and it's important you use the double-M or you'll end up with all kinds of unusual places.

For ACTTION, it's A-C-T-T-I-O-N.org, and you can go to that website. There, information will be about meetings, about all the past papers that have been developed, publications that have appeared, who our partners are, and who's been involved in those different meetings.

So what are the objectives of this particular meeting? Well, in general, the objective is to discuss important considerations and provide suggestions regarding the design,

Page 13

implementation, and evaluation of pragmatic and comparative effectiveness clinical trials of pain treatments. We want to discuss and consider what needs to go into those types of studies and what should someone be thinking about if in fact they're planning on conducting one of those studies.

We want to disseminate these considerations, observations, suggestions, and research agenda by publication in peer-reviewed journals. The only way that we can have an impact -- to use a bad pun, if I will -- is to get information out there. So our goal is to make sure that we get enough information developed so that we are able to develop a manuscript and publish that manuscript.

All of you will be asked, if you participate in the meeting, if you want to be considered. When manuscript drafts are developed, they're circulated. You'll have opportunities to comment on those. The lead author will integrate that information, submit these, and pass them around until we come to a final consensus in the manuscript.

Page 14

Important, when we get to that manuscript, as you can tell by the number of people that are participating in this meeting, that's a lot of authors. If in fact everyone doesn't efficiently respond, it will take a very long time before we can get these manuscripts out. So we encourage you, please, if in fact you choose to be an author, and you read the draft of the manuscript, and you have comments on those, to really provide us with that information as quickly as you can, within a reasonable time frame of 1 to 2 weeks if at all possible.

So that's what we're going to be doing for today, and now we're going to be introducing who the moderators will be for morning session.

I would be remiss if I didn't thank some people who really made this meeting possible. This is the first virtual meeting that we've had for IMMPACT, and we could not have done this without the expert assistance we've had from Jana Hatton [ph], who is the video production person/producer; Carlos Rodriguez, who has been the

Page 15

technical person; and most of all and important is to Valorie Thompson, who has really been the person who's been able to pull all of these things together into a format that we are actually able to use. We really are counting on their support throughout, and I hope you'll be there with us. As I mentioned, this is our first one.

What we'd like to do now is to introduce you to two moderators for the first session. Dr. Robert Edwards -- Rob Edwards is an associate professor at Brigham and Women's Hospital and Harvard -- and Dr. Daniel Cherkin.

Dr. Cherkin was a scientific investigator at the Kaiser Permanente Health Research Institute. Dr. Cherkin is emeritus, but we've been able to attract him to come to help us and to provide us with his insights, knowledge, and wisdom as we move this forward. So let me turn this over to Dr. Edwards and Dr. Cherkin.

DR. EDWARDS: Thanks very much, Dennis, for a terrific overview and summary, and thanks very much to IMMPACT for organizing this exciting

Page 16

meeting. I believe my co-moderator, Dr. Cherkin, will be introducing the next speaker, which I think we can move to now.

DR. CHERKIN: Thank you, Rob, for stepping in there at this lapse there.

I'm happy to present the first speaker, Dr. Lynn DeBar, who is senior investigator at Kaiser Permanente Washington Health Research Institute in Seattle, and who has a master's degree in epidemiology and biostatistics and a doctorate in clinical and health psychology. Dr. DeBar has a unique experience in research that's pragmatic, having worked in the VA for a number of years, and for more than the past two decades, within the Kaiser system. So I'd like to ask Dr. DeBar to kick off the meeting with the first presentation.

Presentation - Lynn DeBar

DR. DeBAR: Thanks for that introduction and the opportunity to participate in this meeting. I wanted to start out just by really thanking both IMMPACT and the NIH Health Care Systems Collaboratory -- it shouldn't be on the same line

Page 17

there -- and infinite thanks to Valorie and Carlos for their patience in what has been a couple days recording ordeal.

The work I'm going to talk about today really reflects work we did as part of the Health Care Systems Collaboratory and also some PCORI-sponsored work. We have been really lucky to have a number of current HEAL initiatives, where we've had to employ a lot of these designs and really wrestle with some of these issues. VA has really pioneered the way for a lot of this, and finally I really wanted to thank our fantastic research teams and collaborators who really deserve a lot of credit in what I'm talking about today.

To start with some definitions and comparisons, there are a variety of ways that comparative effectiveness and pragmatic trials have been written about, and what I wanted to start with is just to recognize where the commonalities are. Comparative effectiveness -- actually, both of these have been done for many, many years, but in more recent years there's been an influx of funding

Page 18

and really some focused attention that have allowed us to make some methodological advances.

For comparative effectiveness, ARA really ushered in a lot of funding and focus on this kind of work and the Affordable Care Act. With that, the funding of the patients that are at Outcomes Research Institute really gave us a leg up in doing this work.

When you look at those Venn diagrams in the center, what you can see is comparative effectiveness includes randomized clinical trials, but it also has a host of really very important methodological advances in observational studies, which maybe we'll touch on during the meeting.

The question that's quite vocal and maybe somewhat unique to comparative effectiveness research is really the tailoring to say what works best for whom. Pragmatic trials -- again, things that that could be considered pragmatic trials -- have been done for a number of years. But the illustration off to your upper right, really showing that funnel from basic research all

Page 19

the way to, really, use within the clinical care settings, there was a recognition several years back that that is a 17-year odyssey, on average, from publication to real application of findings. So really, the impetus in funding and focus on pragmatic trials at the National Institute of Health was to try to really accelerate that pipeline.

The availability of big data, and with that I would say the adoption, in a widespread way, of electronic health records for healthcare delivery systems, really allowed a lot of this work to happen.

Part of the hope with pragmatic trials was that by having these additional tools, we could do bigger, faster, cheaper trials, and really look at a variety of factors that we haven't been able to in smaller trials. Then finally, implementation research on the bottom, both of these approaches intersect with implementation research, and I'm sure we'll learn about that in the meeting.

To start with comparative effectiveness

Page 20

research, the hallmarks of CER are really the head-to-head comparison of two or more medical treatments or services. Those might be like treatments such as illustrated on the left of different medications. They may be also treatments for the same condition that really have a lot of uniqueness to them, so it really covers a variety of different situations.

Importantly, the purpose of this, really, is to assist stakeholders' various strengths, so consumers, clinicians, purchasers, policymakers, in making informed decisions. Then finally, I emphasized this earlier, the ability and the focus on looking at heterogeneity of treatment effect. What works well for whom and how do we tailor treatment is really fundamental to CER.

Pragmatic clinical trials overlap in several important ways, but the motivation here is in some ways a little bit distinct, and that is motivated by relevance and efficiency. These are large efficient studies that are designed to be conducted in real-world delivery systems. In the circles on

Page 21

the bottom, you can see some of the hallmark features.

The study questions and outcomes are shaped by clinicians, policymakers, and patients of what are the really critical questions for these folks in the front lines of care. They're designed, generally, to be deeply embedded in everyday clinical practice. They emphasize interventions that are practical and sustainable in routine clinical workflow. Data in these really rely on electronic health records whenever possible; and finally, a target in the population with greatest needs, so very broad populations with few exclusions.

Just quickly, these are by no means exhaustive but are some of the key parameters that are considered as we're designing trials and just a brief comparison of how comparative effectiveness trials and pragmatic trials really differ in many important ways from our more conventional explanatory RCTs.

As you look at the study question,

Page 22

increasingly, you have stakeholders really at the table in defining and shaping what the study question is. In terms of the question answered, we're really looking in conventional trials at treatments under more ideal conditions where we can really maximize internal validity and looking at what should we believe versus what should we do with a pragmatic trial, and I would suggest CER is really the middle of that.

In terms of comparisons, I'd already mentioned that comparative effectiveness really are two or more active treatments, whereas often for pragmatic trials, it's been a real-world alternative treatment versus usual care, although it can be head-to-head trials there. But those are both contrasted with treatment versus no treatment or versus placebo, more common in conventional trials.

Data collection, interestingly, for true and at the end of the continuum pragmatic trials, relies more on point-of-care clinicians and electronic health records and registries for

Page 23

collection of that information, whereas conventional trials are really outside the care delivery and more of a hybrid, I would suggest, in CER.

Outcomes, again really important emphasis on relevancy to clinicians, patients, decision-makers, with more of a reliance on EHR and secondary data in pragmatic trials. Then finally with treatment adherence, much more rigorous enforcement with conventional trials, and that really is a continuum for CER and pragmatic trials.

As I'm talking in the next few minutes, I'm going to pull along an example of a recent pragmatic trial we did as part of the NIH Health Care Systems Collaboratory called PPACT, and I wanted to say just a couple of words about that so it's familiar as I refer back to it.

This was a trial that we did in three regions of Kaiser Permanente: the Pacific Northwest, in Georgia, and Hawaii. It really was a trial to work with patients who had a variety of kinds of chronic pain conditions, and all were

Page 24

patients on long-term opioid treatment who are really difficult for the primary care providers to treat.

What we did is we took what has been used by many of you and has been looked at for many years as a multidisciplinary intervention. What was unique about that was that it was embedded into primary care and we did it in a very pragmatic way. So this was a cluster randomized trial, clustered at the level of the primary care provider. The intervention I just talked about. The outcomes were very limited in terms of patient-reported outcomes and very reliant on the electronic health record.

In shifting to talk about the fundamental design issues that we need to grapple with, both for pragmatic trials and CER, I wanted to run through some broad overall design issues and then spend some time on talking about patient-reported outcomes and also the importance of stakeholders and the qualitative end of the spectrum.

Any of you that have been involved in

Page 25

pragmatic trials have probably seen this PRECIS wheel. Essentially, this represents a continuum between pragmatic and explanatory trials, and these are scaled from things that if they're right around the center are due to be quite explanatory, and as you get further to the outside of that wheel, it is more pragmatic. These are several important dimensions of the way trials are designed, and they allow an evaluation of how pragmatic a trial is.

This was a publication that was put out early in the first round of the NIH Collaboratory pragmatic trials of which PPACT, that I just introduced you to, was one of these. The reason that I'm showing this is just to illustrate that there is enormous variation in how pragmatic various elements of design and approach are for different kinds of trials.

You can see PPACT there in the upper center, and it looks somewhat like a deflated tire. So we're not way out on the spokes, and I think we'll talk quite a bit more about this. But I think there are reasons that sometimes we want to

Page 26

emphasize things like adherence in delivery of something like a behavioral treatment model, whereas we can be quite pragmatic in the outcomes we look at, and analysis, and so forth.

LIRE is another study that my colleague Jerry Jarvik did, looking at inserting findings from diagnostic images for folks with back pain in the EHR and feeding that feedback back to primary care providers. You'll note with that one it's much more of an inflated cycle, so easier to get into the extremes.

In thinking about how you think about explanatory and pragmatic features, I'd suggest it really is a balancing act. I just noted that an evenly inflated tire may not be possible or the goal. Sometimes trying to be highly relevant to decision-makers, whether those are patients trying to decide on treatment or policymakers, that can come at the expense of trial efficiencies.

For example, patient-reported outcomes might be very important for patients and the decisions they make, but those things might be add-ons to

Page 27

conducting a simple and inexpensive trial.

So there are inevitable external/internal validity trade-offs, and things like relevance, feasibility, timeliness are things to consider in addition to those kind of parameters and dimensions of the PRECIS wheel. A lot of times these drivers are really not primarily scientific, but they are things that you have to do in conjunction with the delivery system that you're working in.

To shift gears and talk about pragmatic trial designs, some of the common ones, and some of the benefits and liabilities, particularly early pragmatic trials often were cluster randomized. I introduced PPACT and noted that it was a cluster randomized trial and that we clustered at the level of the PCP. If you look at that diagram on the upper right, you can see those bottom squares denoting patients for example, and that middle rung would be the primary care providers and that upper rung, the clinic.

If you think about when and why one might cluster, fundamentally it is the level of the

Page 28

intervention that you're doing; are you primarily touching the patient, in which case you might be able to do individual randomization. In our case, in PPACT, we really fed a lot of information back to the primary care provider, did a lot of training for how to interact with their patients around adoption of behavioral skills training, and broader ways of considering chronic pains.

We were really intervening at the primary care provider level, but the most obvious place sometimes is to do things at the clinical level if you're going in and you're really changing practice at a clinic level.

The disadvantage of that, we had started with the idea of clustering at a clinic level, but when we looked at clinics across the places I was showing you, Hawaii, Georgia, the Pacific Northwest, some of those clinics were small.

Hawaii has literally single-room clinics and some of them were immensely large, and trying to reach some sort of balance across those clinics was just too hard to achieve, which is part of the

Page 29

reason we went down to the primary care provider level. But in doing so, there were some things that were lost by our ability to really go into a clinic, and the logistics are much simpler if you can make changes at a clinic level and really bring everybody along with you. So there's, again, both scientific and pragmatic considerations.

The other thing to know about clustering is that the patients, in this case that are clustered under the primary care provider, are going to share variants in a way that needs to be accounted for analytically, so it increases the complexity and also increases the end.

In practice, a lot of times the size of those clusters can be quite variable. It really works best when you've got many clusters and maybe fewer patients in them, rather than to have a few primary care providers with a lot of patients or a few very large clinics; so all things I think we'll touch on.

Stepped wedge is another design. I introduced Jerry Jarvik's LIRE's study, where they

Page 30

inserted the findings from diagnostic imaging to really educate primary care providers on what those meant and maybe to discourage inappropriate referrals for surgery and so forth.

When Jerry did that study, he went through clinics and he added this feature to the electronic health record in the various healthcare systems and clinics he was working on, and that was stepped over time so that by the end of the recruitment period, all of those clinics had that intervention alive and going. The advantage of that is that everybody ultimately gets the intervention, and if successful, you can really sustain it because it's already rolled out.

The thing to consider -- and I think a number of colleagues who have done stepped wedge has found this -- is that if there are any kind of confounds over that time, it can really wreak havoc with trying to make sense of it. In Jerry's case, there were changes in the electronic health record platforms that needed to be adapted to that really created some challenge there. An external kind of

Page 31

contextual change that we've all experienced is COVID and how that might affect files, so this kind of design is very vulnerable to those kinds of things.

Finally, I wanted to just introduce a design that's not as well known but one that if you have secondary data can have a number of advantages. I mentioned that what we're trying to do in comparative effectiveness and pragmatic trials is to maximize external validity. One of the things that I think many of us have experienced in pain trials is that those patients that primary care providers and others think might most benefit sometimes are the most reluctant to enroll in trials.

When we looked at our enrollment in PPACT -- and colleagues in the VA, several of you who are here, I think all of us sometimes top out between 10 and 20 percent of eligible patients that actually enroll. An encouragement or a Zelen design actually solves for that in some ways. This is a design where everybody who is eligible -- and

Page 32

you need secondary data from the electronic record to define eligibility -- are automatically enrolled in the trial and followed over the course, and then half of those people are randomly selected and invited to take part in the intervention.

In these circumstances, we assume that a good portion of those people will decline, but we will continue to follow them with intent-to-treat analyses. It also gives the option that those people who might not immediately warm up to participation but whom you can continue to approach -- just as we do in clinical care when people might not be ready to enroll in a particular intervention -- and you can do that at any point, say, in a year that you're following those patients. So it's quite an intriguing design. It's used increasingly with very stigmatized conditions like substance-abuse disorders, but we could also consider it for pain.

I'm going to shift and talk about patient-reported outcomes, really an important piece of pain-related research, but often things

Page 33

that aren't baked in, in a regular and frequent way, into our clinical workflows in everyday clinical care. When we started PPACT -- as I mentioned, this was focused on people who are all on long-term opioid treatment -- we thought, really, we had set things up well because in the Kaiser regions that we were working in, in theory, those people were all on opioid therapy plans, which included the regular administration of the briefing inventory.

So if you look at that right-hand side, that shows all the elements of the opioid treatment plans, and I've highlighted or boxed and read that that included, or was in theory included, quarterly administrations of the briefing inventory.

What we found when we started the preliminary UG3 year for that trial was, in fact, very often those assessments were not happening. When we really did some of the formative work with our primary care providers, what we discovered is that 12 items was way too long to be administering in everyday care. That also started with the pain

Page 34

intensity items, which they said that's really giving the wrong message; we really want to be focusing on functioning. They did not find this to be an easy tool to use with patients.

So part of what we did in preparation for launching the main trial was to shift the entire Kaiser system over to the PEG, which now is used much more broadly, so a 3-item, smaller set of the BPI. We also used sleep because that was one of the functional items that our PCPs cared a lot about, and there had been some validation of a 4-item version of the PEG that included that. But we found that much more actionable, and I think as we talk about PROs, it's really important to keep some of these pragmatics in mind.

The other thing that we discovered that primary care providers really did not like was the idea of doing something that the PROMIS does well, where the group of items can be variant. They really wanted the same items over time so they could look at change and talk about that very concretely with the patients.

Page 35

Then on the bottom, this is just to indicate you also have to do this in a context of some kind of population management support tools. So the Epic reporting workbench does this, and this can bring up for clinicians who is due for measurements of these kind. But even with those tools, it's important to keep in mind that there are going to be probably some biases in the information that's collected. So patients with more severe disorders often come in more frequently. We've got more PROs measures in them, and that really needs to be [indiscernible].

So what did this really look like in PPACT? Well, we did rely some on what was collected and we did shift the primary tool that was being used to the PEG with the sleep item. But we also had to augment what was happening in clinical care, and we did that by using the patient health records, so pushing out through Epic, in a quarterly basis, this to patients by secure email. If they didn't respond, we used the interactive voice response system and recorded the responses that way, and

Page 36

that meant that only a minority, so roughly a quarter of the population, actually needed to be reached with a live person. I think that's the other thing about these kinds of trials, is we need to automate whatever we can so we can spend our resources very wisely.

So shifting gears now and talking a little bit more on the gray side of the aisle and a little bit more focused on CER, as you look there on the left, those are the fundamental questions PCORI has posed as being critical for the clinical trials and, frankly, all of the research that they support. So it's really based on patients and providers being able to make decisions about what is best for them under what circumstances.

As you can consider those questions, none of those questions can really be stated as a null hypothesis. We are not looking about is this different from not doing anything. It really is a comparative question, and those kinds of comparative questions of two active interventions take a different analytic approach. So I'm hoping

Page 37

in these two days that we talk more about some of the Bayesian approaches that really undergird this, undergird things like adapted trials that really take into consideration what are the prior things that we know that are built into our analytic models.

As I've mentioned, heterogeneity of treatment effect is really the quantitative means of looking at who does best with what, but I'd also suggest that you need to look at that in a qualitative way. So as it's noted on the bottom, not everything that can be counted counts and not everything that counts can be counted.

So that really leads us to talking about stakeholders and their roles in these kinds of trials and a couple dimensions to think about with this. Your customers in performing these kinds of trials are many and varied. I've put in the smallest font, funders and other researchers because even though that's where we often start and we gear how we construct our designs and the hoops we go through, it really needs to be designed with

Page 38

these other stakeholders in mind.

I would say that, really, this can overlap, but there really are two kinds of stakeholders that you engage with through these processes. One is who uses this evidence and makes decisions? When do they make those decisions? How do they make those decisions? Those are all important features in how you design the trials. Then equally as important, when you're conducting these trials, what do you need to know? Who do you need to work with in order to really be able to do these well when they're deeply embedded in clinical workflow?

The clinical workflow alignment is key. The healthcare system partners are often the ones delivering the intervention and the team designing the study, and this is definitely a team sport to be doing this, but they're often the smaller circle, as I've shown here.

The last thing I want to emphasize here is the win and why of stakeholder engagement. This is really cradle to grave, so to speak. PCORI has really brought this home, but really engaging right

Page 39

from the start, what are the relevant questions of high priority? In lighter blue there is that you want to really have these folks working with you all along the way. Things are often very volatile when you're doing these kinds of projects, and you need those champions, and you need to problem solve in the moment; then finally, to help with some of the implementation and really get the findings out to the important audiences.

This is just a note that when one does do qualitative work in these kinds of trials, there are less conventional tools that can be very helpful, and I'm happy to talk more about this in discussion. We use things like weekly journaling. We did postcards back and forth with our stakeholders so that they were really informed as we went. So it was really a two-way street. Rapid assessment approaches can be extremely helpful. These things aren't used in isolation from more traditional qualitative methods but are really critical to this approach.

In closing, what I would suggest is what I

Page 40

hope we're talking about over the course of these two days are really the things here in the lower-right quadrant, emphasizing external validity and probably dealing with more wicked problems, very complicated patients, and complicated analytics to really untangle this, but I am enormously impressed that IMMPACT and ACTTION are taking this on.

Things to consider, many of which I touched on today, are to really carefully consider design and measurement in order to maximize these things; generalizability, who you're working with, the settings in which you're working in; trying to embed this within the everyday clinical workflow; and designing for sustainability. Many of these things, as I've noted, are questions that are of keen interest to our delivery system partners, and when that's the case, there's really an underbelly that you need to plan for.

When I talked about PPACT, we were doing that, really, at the height of the time that people were trying to make the pivot from long-term opioid

Page 41

treatment being the modal kind of treatment, so there was all kinds of QI going on at the same time. There was all kinds of unpredictable change. We really needed to do surveillance to see what was happening in usual care.

I would say this requires partnership in what I would call a vertical manner, meaning that you need everybody from the C-suite to all of the directors in the clinics, to what I think we undervalued initially, the medical assistants who sometimes were closest to these patients and could achieve the most change; so to pay attention and to continue to partner along that entire continuum.

Then finally, I think this is something that this group can really move the needle on, trying to be brutally realistic about what is needed for routine collection of patient-reported outcomes. What are the tools that are really going to be acceptable and feasible in real-world care and how can we get those things used in a regular way that would allow us to do things like use Zelen designs and rely on EHR data in a way that we have more

Page 42

confidence?

So I leave you with just a couple of references that are, I think, key for thinking about CER work. PCORI has really brought that along ways in the years that it's been at the table, and increasingly there have been partnerships in some of the grand rounds. The Health Care Systems Collaboratory has joined with the PCORnet, so I would encourage people to look at these, and I look forward to the rest of the meeting. Thanks so much.

DR. CHERKIN: Well, thank you, Lynn, for that overview of very complex issues facing those of us who are trying to advance science through these new approaches. We have time for a couple of questions now.

One is a question that says, "In all the trials that you have done, which one do you think is most beneficial for the patients? Which one provides the best way to aid in the treatment of pain and function for a person with pain?"

DR. DeBAR: Boy, that's a big question and a

Page 43

hard one to tackle. I think one of the things about pragmatic trials and comparative effectiveness is it allows us, in the best instances, to best mimic the kinds of conditions and flows that show up in the everyday clinical care.

I made mention of a Zelen design, an encouragement design, in one of the slides, where it really takes into consideration everybody who has a particular challenge. In that case, it's a trial that we're looking at opioid-use disorder and concomitant mental health disorders, but it could be something we use for pain. I think one of the most limiting things that we discovered in our last completed trials was that we just didn't have time to get folks into the trial who were somewhat ambivalent about care at the front end, and I think that that's more common than not.

So if we can think about employing designs that really allow for a period of time where people may consider being enrolled but it really takes some time to get them on board, I think it's a much

Page 44

closer replication of what happens in real-world care. I think we haven't talked as much about how stigmatizing pain and related conditions can be and that it takes some things for people to get over the hump of really wanting to participate, particularly in the kinds of non-pharmacotherapy, self-management types of interventions that many of us have championed and are involved in.

So hopefully that answers that. I hope I have more opportunity to talk about those kinds of designs. I also didn't have a chance in that short segment to talk about adaptive designs, which I think really help us to move with the information that we're getting from trials. Rather than have an extended trial that may by the end be following rather than leading what's happening in the field, we can make adaptations as we go.

DR. CHERKIN: Thank you, Lynn. Yes, there will be more time for more discussion of this and other questions.

One other maybe simpler question, "You pointed out the spectrum between the more

Page 45

scientifically rigorous classical efficacy trials and pragmatic trials at another end of the spectrum. Is it appropriate to conduct a pragmatic clinical trial that evaluates a treatment that has not yet been convincingly found to be efficacious in a more rigorous trial?"

DR. DeBAR: Yes. That's such a good question. I hope we talk more about that. I would argue that there are some things that can really only be done in ways in which they're embedded deeply in our delivery system, even if we have incomplete knowledge. Again, I'm going to reference our current NIMH trial, where we're looking at a number of comorbid conditions that show up with opioid-use disorder.

In order to try to do something like that, you have to collaborate with delivery systems and have a large enough sample, and you've got to be doing it in ways -- and I spoke a little bit about this -- where you're using the kind of tools that are really available in our care delivery system.

We talk about hybrid designs where there is

Page 46

effectiveness/implementation, but there are instances where I think there are ways, at least, in which we're putting various components together that we haven't tested well. But if we don't do those in the settings in which this is conducted, they're hard trials to run. So I think this merits a lot more discussion because it's a hard space to really be sure you're simultaneously balancing the things about external validity and really the rigor to be able to interpret what you're doing.

DR. CHERKIN: Okay. Thank you, Lynn.

We'll move now to the next speaker, Dr. Scott Evans, who is a professor and the founding chair of the Department of Biostatistics and Bioinformatics at The George Washington University. He has an interest in design, monitoring, analysis, and reporting and education in clinical trials and diagnostic studies. Beyond his expertise in biostatistics, he also has become very knowledgeable about the study of infectious disease.

Scott?

Page 47

Presentation - Scott Evans

DR. EVANS: Let me begin by thanking Professor Dworkin and IMMPACT for organizing this important meeting and for giving me the opportunity to be part of it. I'm going to talk to you today about pragmatism from a biostatistician perspective. The further I go in my career, the more I'm able to recognize distinctions between research questions that may be subtle on the surface but nonetheless importantly different, and I think there's lots of room for us to improve on finding and answering the most important questions.

I'm going to begin by telling a quick story. Several years ago I had a leak in the roof of my house, and it created this water bubble in the side of my wall. It was really something I had never seen before. In addition to getting a new roof, I had to re-paper the wall. I had asked my neighbor who'd recently papered a similar size room in his house how much paper did you buy, and he replied, "Six rolls, six rolls of paper." But upon finishing papering of the wall, I'd only used

Page 48

4 rolls, and I told my neighbor I had 2 rolls left and what happened. He said, "Oh that happened to you, too."

Now, this was a lesson that I learned, that I actually asked the wrong question. I asked how much paper did you buy rather than how much he used, so perhaps this was my fault. But this is actually a lesson that applies in pragmatic trials, that in many of our traditional trials, we answer perhaps the wrong question or perhaps an important question, but maybe not "the" most important question.

Two things I've learned about traditional clinical trials over the years; first of all, they're rigorously conducted by experts closely adhering to the highest standards of clinical trials, but many times they're essentially useless for helping clinicians make treatment decisions. Perhaps this was said more eloquently by Dave DeMets, a well-known statistician, and Rob Califf, the former FDA commissioner.

This was a publication in JAMA 2011. "Most

Page 49

clinical trials fail to provide the evidence needed to inform medical decision-making. However, the serious implications of this deficit are largely absent from discourse." What they were talking about was that trials are not pragmatic enough. They're not giving us enough information about the effects of interventions as they would be experienced in clinical practice; that we're answering a slightly distorted question in many of our traditional trials.

I want to make a few opening remarks about real-world evidence and real-world data as they correspond to pragmatism. First of all, they might be considered to be associated but they are not the same. Real-world evidence concerns the data source, evidence that we acquire using non-traditional -- at least non-traditional for clinical trials -- sources like electronic health records.

Pragmatism is about the question, and one does not necessarily imply the other. Studies could be very pragmatic without using real-world

Page 50

data and studies that use real-world data may not be very pragmatic either. So if you're interested in answering important questions for clinical practice, then we should be conducting pragmatic studies. To gain the cost and resource efficiencies of using existing data, then we should consider utilizing real-world data.

What might be considered the standard efficacy clinical trials, the typical trial setting involves control over many different factors, and what that does for us is it enhances sensitivity for detecting effects. We may get faster answers because we're able to see those effects a little bit more readily.

Some examples of what I mean by this sort of control, we have selective enrollment criteria into clinical trials. We may use surrogate endpoints rather than clinical outcomes. We limit the use of concomitant therapies, for example, because we think that may confound the outcomes that we observe. But utilizing all that control may potentially limit the relevancy to clinical

Page 51

practice of the results that we find.

On the other hand, if we conduct pragmatic, or sometimes called effectiveness, trials, the purpose here is to inform decisions about clinical practice and policy. One way to think about this is we're going to evaluate strategies of intervention application for treating patients in practice.

There are a lot of imperfections in clinical practice, but those imperfections are part of the game. They're not to be controlled or muted; they're part of the question. This is distinct from answering questions about biology or mechanisms of action, where you might be trying to evaluate what sort of effects might an intervention have if everybody adhered to therapy and was able to tolerate therapy and so forth.

The pragmatic trial is trying to evaluate how well things work under usual conditions rather than under ideal conditions, and the extraneous variation, and noise, and perceived imperfections that occur in clinical practice are not necessarily

Page 52

to be controlled, but they're part of the game. Let them happen. Let them happen the way they happen.

Characteristics of pragmatic trials, we often have diverse and representative populations that enhances generalizability, heterogeneous real-world settings. So many of the traditional trials we do, we enroll patients from sites who are very experienced at trials. They may be very select in certain ways, but on the extreme pragmatic side you try to do it in everyday care, which may be community clinics and so forth. Flexible protocols for how interventions are utilized, adherence, treatment application and so forth, and important patient-centered outcomes would be important for pragmatic trials.

Here's a contrast between the explanatory or efficacy trials versus the pragmatic effectiveness trials with respect to a number of different characteristics. I'll focus on a couple of these characteristics over the next few slides. In the explanatory trial, we're looking at efficacy and

Page 53

can it work under certain conditions. In the pragmatic trial, it's will it work in practice?

If you look at the setting in which they're conducted, explanatory trials may be very well-resourced settings, places that are used to implementing such interventions. Pragmatic trials are going to be much more consistent with normal practice in the real world and variable. Again, participants may be highly selected in explanatory trials. We might exclude patients who are unlikely to comply and that have confounding conditions or complications.

In pragmatic trials, patients are much more representative. You're trying to get a read on what happens to patients as they come into practice. The variation is often minimized in explanatory trials. We standardize the way we measure things. We standardize the way we apply things. But in a pragmatic setting, they're not necessarily standardized and there's extra noise and so forth there.

I'm going to talk about a few of these

Page 54

issues. There was a tool, this PRECIS-2 tool, that looked at a number of different, which might be considered, dimensions of pragmatism: how the primary analysis was done; who's selected to get into the trial in terms of eligibility; how patients are recruited; what the setting is; and so forth.

There are a number of different dimensions by which you might categorize or classify a clinical trial. A trial may be pragmatic in some respects but not in other respects, and there may be gradations to how pragmatic they are. This was a tool to characterize and classify clinical trials in that way.

Some of the most challenging issues for pragmatic trials in pain, I want to talk about a couple of issues that I think are perhaps most challenging for pain trials. First thing is that pragmatic trials, at least the most extreme pragmatic trials, would not include blinding because blinding doesn't occur in clinical practice. That can be a big issue for pain, and

Page 55

I'll talk about that.

Limited control of concomitant medications is another big issue that often happens in pain trials, as we're controlling concomitant therapy, particularly over-the-counter pain medication; and flexible application of the intervention and questions about the quality and completeness, for example, of real-world data that might be used in such trials.

One concern is a lack of blinding. The most extreme pragmatic designs avoid blinding and placebos. They have real-world control groups without necessarily blinding, and this would avoid placebo effects, which can be a concern in pain trials. On the other hand, we know that pain is a very patient-centered and subjective outcome that can be affected by knowledge of the treatment. Knowledge of that treatment can affect adherence. It could potentially result in treatment cross-over or drop-out of a study. So that's a particularly challenging issue in pain trials to be thinking about.

Page 56

I thought I might show one example. This was actually in asthma. This was a four-armed clinical trial for asthma treatment in which three of the arms actually were non-active. Albuterol was the active treatment for asthma, but there was an albuterol placebo. It was a sham acupuncture and then there was one treatment group that just received nothing. So three of these four arms were actually inactive.

Now, when you compared the change in forced expiratory volume, a common outcome in asthma trials, what you saw was a big improvement for the albuterol arm, but the three arms that did not receive active treatment did less well, and clearly albuterol was superior to the other three arms with respect to this change in forced expiratory volume, an objective measure.

But then there was a subjective improvement that was evaluated by patients, and if you looked at that particular outcome, the outcomes for albuterol, the albuterol placebo, and the sham acupuncture, namely the three arms that thought

Page 57

they were being treated, were quite similar, but the one arm that had no intervention and knew that they were not getting any intervention did less well.

So this was basically showing you that knowledge of the treatment was affecting this particular outcome. The subjective versus objective nature of these outcomes kind of played into these results. Certainly in pain trials, thinking carefully about issues like this would certainly be important.

Other concerns about concomitant medication use, pain trials often place strict rules on concomitant therapy use such as over-the-counter pain medications. The primary reason for this is sensitivity to detect effects and that we can try to isolate the effect of the treatment in question that we're trying to isolate. But pragmatic trials place very limited restrictions on concomitant medications because in practice, patients are going to take aspirin, or ibuprofen, or whatever it might be if they're feeling pain. So these are sort of a

Page 58

dichotomy of an important question to be thinking about if pragmatic trials and pain are going to be considered.

What about using real-world data? Well, it could be very challenging to use real-world data for measuring pain outcomes. We know that depending on the type of real-world data, data quality and completeness can be an issue. In clinical practice, there's not necessarily a standardization of measurements and methods for measuring outcomes, pain outcomes or other outcomes. Some standardization is important in this particular area. Even the time of day relative to when you measure pain outcomes and so forth can be particularly important.

There is a publication coming out from the Clinical Trials Transformation Initiative that was evaluating uses of real-world data. They chose, actually, to focus on using real-world data for specific purposes that may be beneficial but particularly low risk. For example, using real-world data for planning eligibility criteria

Page 59

or for enhancing recruitment, maybe you can access registry data to know who has a particular chronic disease that might have pain.

That's fairly low risk to try to help identify who might enter a trial, but if you're going to be using it to measure outcomes on patients, that's higher risk. There may be some hybrid by which you can use real-world data for some purposes, but if you're measuring, say, outcomes on patients, you might want to standardize and make sure that those types of outcomes are more complete and are done in a high-quality fashion.

Some of the general concerns about pragmatic trials is this extraneous variation that is not controlled from either limited restrictions on entry criteria or the use of concomitant therapies may result in a dilution of the treatment effects; that if the noise is bigger than the signal, you won't be able to see the signal. There's been some mixed results about whether that's actually the case. Some research suggests that it really may not dilute the treatment effect as much as you

Page 60

think, but that's one of the main concerns, is that the noise is bigger than the signal, which may create problems in identifying the signal.

The other thing to be aware of is if we conduct trials or try to use real-world data, pragmatic trials often have large N, large sample sizes. But large numbers don't eliminate biases and confounding. They give you more information, but they don't eliminate biases and confounding. So be wary of that or don't mislead ourselves that just because we have a larger sample size, bias and confounding are going to go away.

There are opportunities, I think, in the real-world data setting and getting data through personal devices. We've done trials in which we have been able to collect pain data, for example, through prompts in phones and other personalized devices, so I do think that's a big opportunity. If we could ever really get smart about marrying the research enterprise with the clinical practice enterprise, and being able to use data and research from the clinical practice infrastructure, that

Page 61

would really be valuable.
I wanted to make a couple of comments about some statistical concerns that are a little bit less recognized in my view, and that is that typical approaches to analyses, statistical analyses, are often not very pragmatic. This has really not been recognized in many places, so I'm going to show you a couple of examples of what I mean.
Here's a question. We define analysis populations in clinical trials and efficacy analyses. We use an intention-to-treat population. We do a safety analysis in a safety population. Those two populations are not the same. At the end of the trial, we may try to combine these analyses into something we might term a benefit-risk analysis, but when you step back to whom does this analysis apply, what are we actually estimating?
Well, in statistics, we estimate a parameter, which parameters are characteristic of a population, but here we have outcomes being assessed in different populations, so there isn't

Page 62

even clarity about what population we're talking about. Then when we start thinking about personalized medicine, how do we do personalized medicine if we're not even evaluating what's happening to the patient? There may be correlations between efficacy and safety outcomes, and I'm going to show you an example of that.
Question number 2. Pain is often a symptom of an underlying disease and the effectiveness of pain management can affect the underlying disease outcome. For example, if you were a failure of chemotherapy, maybe a downstream result of failure to control peripheral neuropathy might be associated with that chemotherapy, and in turn the change in chemotherapy affects peripheral neuropathy outcomes.
We have this difficult problem of competing risks, and circularity, and the fact that pain is a symptom of an underlying disease. Yet oftentimes we're trying to separate outcomes of the disease versus outcomes of treating the symptoms when, in fact, they're actually connected and can affect

Page 63

each other, and that's a very difficult problem.
Suppose you measure the duration of a chemotherapy-associated peripheral neuropathy, and you say, well a shorter duration of that peripheral neuropathy is better? But the faster the patient withdraws from chemo, the shorter the duration, or the faster the patient dies, the shorter the duration. So the interpretation of one outcome really needs context of other clinical outcomes for that same patient; yet, we often are analyzing them separately. So we have to be careful about that, and I'll show you another example of this.
We've had some great work on estimands, meaning clarity about what population parameter you really want to estimate, and we finally get people to recognize that if we change the analysis population, that means we're changing the question we're addressing. Here's an example of what I mean.
Suppose we conduct a randomized trial that compares two treatments, A and B, and a trial participant that is assigned to treatment A

Page 64

discontinues A and begins a new intervention C. The participant then experiences an SAE. It's adjudicated as related to C but not to A. So this leads to the belief that safety is not an issue for A; it's C's fault. So we might continue the trial and say, well, I can still do this randomized trial of A versus B because it wasn't A that was the problem.
Now suppose 10 additional trial participants discontinue A, begin treatment C, and experience the SAE. Again, adjudication links the relationship to C but not A, but there's no such events in arm B. Would you volunteer to be randomized into this trial possibly with being randomized to treatment A? Well, if you were on the data monitoring committee, can you allow continued randomization into A?
So the issue here is that adjudication, despite the fact that biologically you might attribute this to treatment C, the strategy of applying A is a problem because there's nothing happening in arm B, and I wouldn't want to be

Page 65

randomized to treatment A under this case. So this is an intention-to-treat issue, that what you're interested in from a pragmatic point of view is the strategy of its application. This is evaluated in a randomized setting. You cannot allow a continued randomization into arm A in a case like this, and we've talked about this in a recent paper in the context of data monitoring committees.

Here's another question. Suppose a loved one is diagnosed with a serious disease and you were selecting the treatment, and you have three treatment options, A, B, and C. Let's suppose for simplicity that there are two outcomes equally important: a treatment success, yes or no, and there's a safety event, yes or no.

So we have two, an efficacy outcome, a safety outcome, and both are binary, and let's suppose they're of similar importance. Now, luckily enough, there was a randomized trial that compared these three treatments, and we'll look at those outcomes because that can help inform whether we would choose A, B, or C.

Page 66

There were 100 participants in each of the three arms. The treatment success rate was 50 percent in A, 50 percent in B, and 50 percent in C. The safety event rate was 30 percent in A and 50 percent in B and C. So which treatment do you choose? They all have the same success rate. A's got the lowest safety event rate. B and C are indistinguishable. You can't even tell the difference between B and C.

So since we're all logical people, we'd probably choose treatment A, but instead of taking the patients in the trial and analyzing the outcomes, which is what we have done to this point -- we took the patients in the trial, the hundred in each arm, and analyzed the outcomes -- what I'd like to do is flip that upside down; take the outcomes in the trial and analyze what happened to the patients.

Now, if you think about that play on words, what that means is that there are four possible patient outcomes. You either get the treatment success, yes or no, with or without the safety event, yes or no. Our culture has been to use these patients to analyze the outcomes, but shouldn't we use the outcomes to analyze the patients?

Page 67

So here's what I'm going to do. We cross-classify and analyze the patients. So it turns out that in treatment A, the treatment success and the safety event were uncorrelated, so there were 35 patients that had the treatment success without the safety event. In arm B., the treatment success and the safety event were positively correlated, so there are zero patients that had the treatment success without the safety event. In C, they were negatively correlated, so there are actually 50 patients who had treatment success without the safety event.

Now, one or two slides ago, we couldn't tell the difference between B and C. If we take the outcomes and analyze the patients, we may see some things that we're not able to see by taking the patients and analyzing the outcomes.

Question 5. A negative trial may not mean that the treatment is good for no one. A positive trial doesn't mean that it works for everyone. How do we identify the subgroup of patients we want to treat? Well, let's return to the example we had. If we had predictive markers for efficacy, then we'd identify these people. If we had predictive markers for safety, then we'd identify these people. But neither of these approaches finds the patients that you actually want to treat.

Page 68

So even thinking about this benefit-risk type of approach, what you're interested in is identifying the patients who have treatment success without the safety event, but by looking at one outcome at a time, we're unable to find those patients. So shouldn't personalized medicine be based on analysis of the patient?

I call this the clinical trial arithmetic. What we do is we compare each outcome and then combine how those outcomes compare. But what we really should be doing is combining outcomes within patient and then comparing how treatment A compares with treatment B. That's what's most relevant for

Page 69

evaluating the effects of treating patients. As my father told me many years ago, the order of operations is important, and the clinical trial of arithmetic has never been quite right.

There are some evolving methods about how to do this. We had written a paper with a colleague, Dean Follmann, using outcomes to analyze patients rather than the patients to analyze the outcomes, and this would be a step towards pragmatism and benefit-risk evaluation. We have some methods called desirability of outcome ranking that try to address things in this way, and the idea is that before we analyze hundreds of thousands of patients, we really have to figure out how to analyze one by combining information within patient and evaluating the patient journey through the trial, sort of a synthesized analysis of the benefits and the harms of the quality of life that has been experienced by the patient.

The way I describe this is there's a map to pragmatism as I see it. Where we are today is when you think about the number of outcomes we have in

Page 70

trials, it's fairly high, but we estimate one effect for each of those outcomes.

Where we'd like to be is pragmatism on the far right there, where the number of outcomes we have is actually reduced and that there's a patient outcome, and that patient outcome composes individual outcomes on efficacy, and safety, and so forth. But then we would estimate, and the path we take to get there is we make progress via composition of the outcomes, characterizing the disease burden for a patient, but then move towards personalized medicine by estimating the effect on that total disease burden, depending on the characteristics of the patient.

I'm going to show you one quick example of this. This is the PROVIDE study. It was a prospective, multicenter, observational evaluation among adult hospitalized patients with MRSA bloodstream infections. The research question was what PK dosing target would be optimal for treatment outcome. So if you think about dosing, well, I want to make sure I get a high enough dose

Page 71

so that there's efficacy, but if I get too high, it's going to be toxic. So it's a benefit-risk problem. How do I look at that?

There were 265 patients in this particular study, and we set up what was called a desirability of outcome ranking. Treatment success without acute kidney injury is the most desirable outcome, the patient has a treatment success and avoids toxicity. The least desirable outcome is at the bottom where the patient dies. But then there are gradations of patient response in between in which they may be getting treatment success but they may have toxicity at various levels.

In this particular outcome, this desirability of outcome ranking had five levels, most desirable at the top and least desirable at the bottom, and intermediate categories where some things go right but not everything goes right.

What we did is look at the distribution of this particular ordinal door outcome by dosing levels, and the vertical axis represents the quintiles of dosing, the highest dose on the top

Page 72

and the lowest dose on the bottom, and the different colors represent the five DOR categories. The bluish on the left is a treatment success without acute kidney injury and the purple on the far right is mortality.

Now, as you go from the low doses to the high doses, what do you actually see? What do patients gain? They don't necessarily gain efficacy because this red area is actually treatment success but with acute kidney injury, so what we're gaining with higher doses is toxicity rather than gaining efficacy, and lower doses in this particular case may be better.

So how might this work in pain trials, you may have pain trials where there might be an outcome, say, with four levels. The most desirable at the top is pain control without toxicity and the least desirable at the bottom, but there might be a few gradations in between. There are a couple of different ways in which to analyze this particular outcome.

One method you see on the far right is to

Page 73

score this like an academic test, and if the patient has the most desirable outcome, they get a score of 100; the least desirable outcome, they get a score of zero. If they're in the intermediate categories, they would get partial credit, and then you could analyze it sort of as a difference of means. There are other methods using ranked-based methods.

My suggestions, I think it's important to address pragmatic questions. I think much of the time we're answering the wrong question. We should be evaluating strategies of application. I think it's a better reflection of the value to society. Pragmatic approaches; analysis populations; intention to treat; some flexibility in treatment application; allowance of concomitant medications; analyzing patients rather than outcomes; and using real-world data for certain things but not others I think can be risky for certain things, but try to retain the rigor.

I do think that randomization and blinding is still important regardless of the fact that the

Page 74

most pragmatic of studies is avoiding blinding and things like that. We've got to be careful about using real-world data for addressing endpoints, I think, because we need to make sure that we've got quality and completeness for outcomes.

So I will end there and say, I have no doubt you will enthusiastically applaud now because you're so relieved that my talk is over. Thank you very much.

DR. CHERKIN: Okay. Well, thank you, Dr. Evans. I'm trying to look over the questions here.

One question says, "Dr. Evans noted the lack of blinding was a concern in pragmatic trials and, as he showed in the asthma trial, that albuterol was superior to, and sham acupuncture, in terms of objective outcomes but was no better in terms of subject of outcomes. For asthma, the objective outcome may be more clinically important than subjective patient reports of improvement. However, in the study of treatments for common, nonspecific pain problems, could one argue that

Page 75

subjective outcomes such as increased function, quality of life, and less pain may be more important or relevant than objective outcomes such as improved range of motion?"

DR. EVANS: Yes. Thank you for your question. I think you're right. I think it's a point well taken that many of the outcomes that we are interested in have a subjective nature to them, and those are certainly important.

My point about the blinding, in the most pragmatic form of trials, they try to avoid blinding because you're trying to mimic what's happening in clinical practice as much as you can, in regular clinical practice. This is why I talked about it. I think it's a difficult area for pain trials because blinding in some ways is so important in traditional pain trials. So this is one area where I think it's a real struggle, and real critical talk is going to be needed in pain trials. But your point is well taken, and I think you have valid points there.

I would note that even objective outcomes

Page 76

are not immune to concerns about a lack of blinding. Patients can selectively drop out and so forth, and that can still affect that for even objective outcomes. But my point was to raise this as an important consideration, or one of the main or biggest issues for pain trials is thinking about the blinding question and wrestling with the pros and cons of that.

DR. CHERKIN: Okay. Thank you.

Rob, I've kind of goofed up my screen. Could you ask the question that you forwarded to me, please?

DR. EDWARDS: Definitely. That's why they have two of us as co-moderators. I'm happy to jump in.

DR. CHERKIN: Thank you.

DR. EDWARDS: Scott, a terrific presentation. We've got a question that just came in from Nat Katz, who also loved your presentation and inquired, "Are you drawing a distinction between pragmatic questions and pragmatic trials?"

DR. EVANS: Yes. I wanted to make sure that

Page 77

some of the terminology -- my point was that some of the terminology, it's not completely well defined and people use it in different ways. What I wanted to make sure that we're able to distinguish is a lot of people will call pragmatic trials those that are using data from non-traditional trial sources, and using health records data, or other sorts of data from the clinical practice infrastructure, but there are a lot of studies there that, although they may use that data, are not necessarily very pragmatic; and there are a fair amount of what you consider to be pragmatic trials addressing pragmatic questions that may not use much of that data.

So I wanted to make sure that we have some language and clarity in our own thinking that there's a distinction between using data from non-traditional sources and potentially from clinical practice and addressing what might be considered pragmatic questions. So I just wanted to make sure there was clarity of thought there.

DR. CHERKIN: Okay. Well, thank you, Rob

Page 78

and Scott. I think, again, like in Dr. DeBar's presentation, this is a very complex area. Pragmatic trials aren't for the faint of heart, and there are a lot of issues to pay attention to. So thank you to some of the pioneers in this area, and now I'm going to turn it over to Rob to introduce the next speaker.

DR. EDWARDS: Terrific. Thanks, Scott, and thanks very much, Dan.

It is a real pleasure to introduce our next speaker, who will be Dr. Ajay Wasan. Ajay is a psychiatrist and a pain physician, and parenthetically an outstanding tennis player as well. He's a professor and vice chair for pain medicine and the co-director at the Center for Innovation in Pain Care at the University of Pittsburgh.

He and his outstanding research team are doing seminal work in areas such as phenotyping of patients with acute and chronic pain; interventional procedures for pain management; pain-related medical informatics; and great work on

Page 79

pain biomarkers using functional neuroimaging and other technologies. I'm very excited to hear his talk on lessons learned from psychiatry, the CATIE and STAR-D trials. So I'll leave it to Ajay to take it away from here.

Presentation - Ajay Wasan

DR. WASAN: Hello. I'm Ajay Wasan, and I'm at the University of Pittsburgh. This over here on the right is actually the flagship building at the University of Pittsburgh. It's the 44-story tall Cathedral of Learning.

What I'm going to cover today is a discussion of a couple of important trials, important comparative effectiveness trials, the CATIE trial, which looked at atypical antipsychotics in Alzheimer's dementia, and it has been very productive and generated over 50 papers, and the STAR-D trial, which is even larger and perhaps even more important than that and more influential, sequential trials of antidepressants in major depression.

I'm coming at this from the perspective of

Page 80

what we can learn from these trials, and I think what we can learn is this idea of balance, of balancing the explanatory and pragmatic components of the trial in order to reach the scientific objectives and understanding what the trade-offs are.

This was an editorial I wrote in 2014 on this issue, Effectiveness versus Efficacy and Explanatory versus Pragmatic. What I argued in this editorial is that there's a sweet spot in pain medicine research, in the field and in comparative effectiveness research, which will yield the most scientifically and clinically impactful findings. The idea is when you're designing any trial, any comparative effectiveness trial, is try to understand what is your sweet spot for what you're trying to achieve scientifically.

This quotation here really summarizes I think the upshot of what I'd like us all to take away from this talk today, which is, "The optimal balance point between the poles of pragmatic and explanatory qualities is where these approaches

Page 81

mutually strengthen each other to create the most robust framework for a clinical trial design."

Dr. Lynn DeBar may have already shown this in her talk, and I apologize for the redundancy, but one very important tool that's out there for helping to balance the pragmatic versus the explanatory is the PRECIS-2 tool. There was a PRECIS-1 version. That was very successful, and it's been validated and refined into PRECIS-2. A nice description of this appears in the BMJ in 2015.

This was a very comprehensive process to arrive at this tool. Eighty international trialists, clinicians, and policymakers contributed to this. What you have are nine different domains over here that represent different aspects of a clinical trial, and you can design each of these domains from the most pragmatic, which is the outer ring, so to speak, of 5, graded a 5, versus the least pragmatic and more explanatory, which is a 1.

You can get a sense in each of these key areas where does your trial lie on that balance

Page 82

point between explanatory and pragmatic. It's very valuable for matching design decisions for how the results of a trial are intended to be used.

Looking at all of this, this is from my own impressions in the literature and my own impressions of these trials of CATIE and STAR-D. What I've summarized here are some key issues, I think, in comparative effectiveness trial design that really cut across multiples of these domains.

For instance, it's particularly important that any interventions that are tested are tested in routine clinical delivery settings when possible. That really speaks to generalizability, which is I think very important, but with attention to minimizing the known confounders. That is, of course, the best approach to causal hypotheses.

It's this issue of what is the magnitude of the unknown confounders when you're doing a study in a more routine clinical delivery setting, and that's what actually stymies and vexes a lot of the clinical trialists and the evidence-based medicine scientists. It's that tension and that conflict

Page 83

that's there, and that is I think a key balance point.

Other key balance points are the inclusion criteria. Do you choose broad or specific? It's important to consider. Duration of tracking the intervention; are you just looking for a signal of the effectiveness? So maybe you only tracked an intervention for two months or are you actually looking for durability, and you want to know what the results are at one year? So important things to keep in mind. Also, what's the specific dose of the intervention? Do you have one dose for all subjects or is there a range of dosing that's allowed for all subjects? Of course that introduces confounders as well that you need to be aware of.

Then randomization and lots of issues over how you randomize, when to randomize, and where to randomize. Do you include crossover randomization or not? All these of course have implications for explanatory versus pragmatic trial design. Outcomes assessment, this is important, too. Do

Page 84

you have very rigorous outcomes assessments for a very specific set of outcomes that you're looking for or are you looking more at global health outcomes and if generalizability, for instance, is the key goal that you want? So again, things that need to be balanced. Data analysis, lots of issues there, for instance, doing intention-to-treat analysis versus only analyzing those who completed treatment.

I'll talk a little bit about the CATIE trial. It stands for Clinical Antipsychotic Trials of Intervention Effectiveness. There were several CATIE trials. There was one for chronic schizophrenia as well, and I won't really go into that. That trial design was similar, but there was no placebo arm. It was only a comparative effectiveness trial of five different antipsychotics, and the upshot found that olanzapine was superior.

What I'm going to talk about more is the CATIE Alzheimer's trial, which looked at olanzapine, quetiapine, and risperidone in low

Page 85

doses -- those are all atypical antipsychotics -- or placebo, in 450 subjects. It was double-blinded and randomized. They did provide 2 doses for each drug, low and high medication doses, so you might start someone on low and then gradually go up. The high dose was not like schizophrenia high. It was quite low actually because, really, just small doses of atypical antipsychotics are used in this patient population.

They had two primary outcomes, and this is very important. It's an important decision they made, and it's unique, and it certainly gets at this issue of the purpose of the trial that you may actually need to create some unique outcomes. The outcome they had is time from initiation treatment to discontinuation of any treatment. The idea is that if a treatment was effective, it would be continued. So if it would help control the agitation and psychosis, it would be effective and it would be continued, but if the treatment was not effective or if there were significant side effects like extrapyramidal symptoms, dizziness, confusion,

Page 86

et cetera, then the treatment would be stopped.

So the time to treatment initiation is really a proxy for treatment failure. That's a primary outcome. The minimal improvement benchmark on the clinical Global Impression of Change scale at 12 weeks, that was the second primary outcome. This is the paper published in the New England Journal, and they had two primary hypotheses that were tested.

I think it's important to appreciate these. One is the three atypical antipsychotic treatment groups taken together will be superior to the placebo group in terms of all-cause treatment discontinuation, and number two, the three antipsychotic treatment groups will be equivalent to each other with respect to response at 12 weeks regardless of subsequent randomizations.

So what they did in this study is first they used fairly rigorous enrollment criteria. Each patient had to have a DSM diagnosis for Alzheimer's and deficits on a mini mental status exam. That's a fairly rigorous approach, but they used a fairly

Page 87

pragmatic approach for some of the specific symptoms that are indications for using antipsychotics.

Psychosis, aggression, and agitation symptoms were determined by the judgment of the treating physician and the patients had to have those for 4 weeks. The prescribing reflected clinical practice, and if a subject was not responsive to medication, that medicine could be discontinued in 2 weeks, and then the subject could move to phase 2.

This is an outline of the design. We're going to focus on phase 1 here, which is each of these 4 arms, placebo versus the 3 antipsychotics. Then say for instance, in 2 weeks if someone failed the first antipsychotic, they could be randomized to one of these other arms, which is the randomization to either of the two antipsychotics they didn't have or citalopram. That's an SSRI. That's fairly well tolerated in the elderly. Then those who got placebo would be randomized actually into the 4 arms, including citalopram, so without

Page 88

placebo.

One thing to think about is this is a fairly sophisticated adapter trial design, and historically we have not had very many adapter trials in our field of pain, and that's something for us to think about. If we designed a big adapter trial, we're sort of pushing the envelope. Is our field really ready for that? Is the science there to support it? So again, things to think about.

This is a summary of the main outcomes, and it's very interesting here. Over here on the left, this is the time to discontinuation. It's a Kaplan-Meier curve, and this is discontinuation for any reason, so lack of effectiveness or for too many side effects. You can see that the drop-off rates, they all essentially are the same. There's no significant difference between them. But if you split it up into discontinuations related to lack of effectiveness -- that's here -- there was a clear separation between all three of the antipsychotics and placebo, which is this orange

Page 89

line at the bottom, and it really appears here that olanzapine and quetiapine are performing better than risperidone.

Then if you look at the discontinuation for the reason of adverse effects, you see the opposite Kaplan-Meier curve. Of course you'd expect those on placebo that have the least side effects and that gets discontinued least often, while the other 3 antipsychotics get discontinued more. In this case, olanzapine was discontinued at the greatest rate with the quetiapine and risperidone somewhere in the middle.

When you look at this second outcome, the Global Clinical Impression of Change, these are some of the comparisons, similar; and there was a trend towards the antipsychotics producing more impression of change, more global improvement than the placebo, but that was not significant.

So some key lessons here that come out of this. Looking at a combined primary outcome of efficacy plus adverse effects is very pragmatic. It may be something very useful to consider for a

Page 90

pain trial, but it would be important to distinguish between the effectiveness outcomes and the adverse effects outcomes.

Another thing is would the assay sensitivity of this trial be improved if the primary outcomes were continuous in some fashion? It's kind of hard to do for what the aims were of this study but something to think about for pain trials.

Also, when you're looking at effectiveness studies, it seems that re-randomizing treatment failures is important, and that's an adaptive component. That's a version of a smart design, and some of the other speakers might be speaking in more detail about what this is. A smart design is a fairly sophisticated adapter trial approach. When I'm talking about re-randomizing treatment failures, really it's just step one in a smart design.

We'll move on to STAR-D. That stands for Sequence Treatment Alternatives to Relieve Depression. It's a much more complicated study, so what I'm going to present really is a general

Page 91

overview about what I think is more important about this study. It was designed to assess the effectiveness of treatments in generalizable samples and ensure that the delivery of the treatments was adequate at the same time.

The study tried to look at the symptomatic outcomes for outpatients with nonpsychotic major depressive disorder who treated initially with citalopram. That was step 1. That was phase 1 in the trial. The question fundamentally of the trial is what you do next if citalopram fails. So if 1 SSR fails, what do you do next?

Having phase 1 being citalopram only, establish treatment remitters versus non-remitters. That was the key. Remitters means that almost all the symptoms of depression go away. They're trying to identify the treatment-resistant subgroup, and just to note, that's a little bit different than treatment-resistant depression because technically that's defined as failing two antidepressants, not one.

What they did in level 1, they gave an

Page 92

adequate dose of citalopram for an adequate treatment trial to look at this efficacy, and this was done in routine clinical settings in outpatient psychiatry practices and primary care. The primary outcome was remission and depression symptoms, and that's also a unique primary outcome. That's almost the absence of depression symptoms versus response.

Typically in antidepressant trials, you define response as being a 50 percent improvement in symptoms since that is very clinically meaningful, but they looked at remission here. To give you some sense of what difference that might include, in randomized trials of antidepressants, typically you'd have a 20 to 30 percent remission rate and a 35 to 45 response rate.

So it's something for an adapter trial to consider, which is what really should be the primary pain outcome? Should we think about things a little bit differently in adapter trials, in these comparative effectiveness trials, versus a classic randomized-controlled trial for pain

Page 93

outcome?
Some of the key features, STAR-D had 4,000 patients. It was very broad entry criteria, nonpsychotic major depression diagnosed by DSM checklist and also having high symptom scores on the HAM-D, which is the Hamilton Depression Symptom Rating Scale. It's been noted in some editorials that actually many of the patients in STAR-D would not have qualified for FDA phase 3 antidepressant trials, so this gets at the increased generalizability that I mentioned before.
The dosing was flexible but rigorous. Citalopram actually had a pretty high dose range. We actually don't use 60 milligrams anymore because of some cardiac issues that can occur. The patient could be increased every 2 to 4 weeks. The treatments were unblinded. They were prescribed by a psychiatrist or a primary care physician. It was measurement-based care, so the decisions by the treating physician for whether to increase the dose or determine that the patient was a treatment remitter or non-remitter were used by symptom

Page 94

checklists.
There was some degree of patient choice for which treatments they could receive. Patients, if they got randomized, for instance, to a cognitive therapy arm, they can choose to reject that and go into a medication arm. This is particularly important. I think it's an important lesson that we really should include patient choice as an arm. It's revolutionary, but it's important, and there's a precedent for doing this. In the STAR-D trials, I think patient choice is crucial particularly when you're thinking about multimodal pain trials and how to optimize multimodal care and do comparative effectiveness around those issues, so I would keep that in mind.
They had 12 to 14 weeks per treatment, per treatment level, and I will show on the next slide what I mean by that and when you follow up. One consequence of the patient choice was that those who were randomized to cognitive therapy, only a much smaller percentage chose to stay in that arm -- no offense to all the psychologists here who

Page 95

are listening -- so they could not actually compare the cognitive therapy to the other medications.
This is a summary of the study flow. Initial treatment, as I mentioned, is citalopram, then patients could be switched to a couple different antidepressants, bupropion cognitive therapy as I mentioned; sertraline; venlafaxine; or they could have an augmentation strategy, which means that you add something to the citalopram. Even if the citalopram is not working, you could add something to it, so bupropion, buspirone, or cognitive therapy.
Then level 3 is if they failed those, then they can be switched to mirtazepine or nortriptyline, or again they could have something augmented. So you could have something added to the other drugs they're on. It could be lithium or it could be a thyroid hormone. Number 4 was actually switching to tranylcypromine, which is Parnate, which is an MAOI, an older antidepressant that has more significant side effects and dietary considerations, or mirtazepine combined with

Page 96

venlafaxine.
These are some of the main studies that are published in the New England Journal of Medicine. Over here are the outcomes of the switch to the different antidepressants. Basically, it didn't matter what you switched to, there was a consistent rate of increased remission. And that's really the wrong word, meaning that it's better to say that more subjects responded.
So those who had not responded to citalopram got switched, and then there was a meaningful percentage of people who remitted when a different antidepressant was used, but it didn't matter which antidepressant was used; and similar outcomes with augmentation. For the augmentation, augmentation worked, but it didn't matter if you had buspirone or sustained-release bupropion, both of those increased the number of remitters.
Some important study takeaways -- and I'll be finishing up; only a couple slides left -- level 1 citalopram achieved 33 percent, who on that in phase 1 achieved remission. Some of the

Page 97

important takeaways were, for all the levels, if there's a partial benefit at 6 weeks, it's important to raise the dose, and you would definitely get more treatment remitters then.

The outcomes of switching to a different antidepressant or augmentation were no different between the treatments, and with this process of persistent and vigorous treatments, most patients actually entered remission. These are the numbers of how it works out. If you go through all the steps, up to 67 percent of all subjects total actually entered remissions, assuming that they stayed in treatment.

So paying attention to attrition and the different phases in relapse is particularly important, and the remission rates did drop off per stage. So by the time you got to the fourth stage, of those who had already been treatment resistant to all these other treatments, only 15 percent did respond with remission. But that's important because some of these patients have significant symptoms like suicidality, and if you can decrease

Page 98

and eliminate that symptom, that's very, very important and meaningful.

Seven weeks was the average time needed to achieve remission, and interestingly, it's important to know in a multiphase study that all the treatments can be compared to each other. It's hard to explain, but due to different considerations in the trial, it turned out that the augmentation arm could not be compared to the antidepressant med switching arm. So it's just something to be aware of that in multiphase trials, they were able to compare.

Over here on the left is just to give you a sense of the number of subjects in each of the different cells. You can see it's fairly complicated how the randomization goes. You look here at the bottom, by the time you get to this level 3 switching and level 4, essentially there's just not enough patients that you can even analyze, so keep that in mind.

Over here is just a graph of the attrition rates and relapse rates. I would say it's an

Page 99

important lesson for our pain studies that we want to think about pain flares. How do we model pain flares in the studies? How do we anticipate them? Do we think of pain flares during the period of follow-up and treatment to be outcomes or are they confounders? They're both, obviously. What do we do with those is particularly important, and the STAR-D trial has shown that.

There has also been a number of downstream reanalyses of the STAR-D data. Most recently, for instance, there's some predictive modeling using machine learning techniques to understand who responds and doesn't respond. For many of these studies and for many of the editorials that have been written, these are some of the key takeaways that popped out to me, that it would be important to understand the characteristics of early responders.

Several of these characteristics had already been identified in STAR-D, such as higher physical health ratings of baseline and lower anxiety at baseline. Certainly for pain trials, we want to

Page 100

think about how do we identify early responders, and do we actually consider re-randomizing at 2 two weeks if there is no response? That's something that would also be fairly novel, but it can also be scientifically important.

Other things in terms of generalizability, thinking about where the treatments are administered, different clinical sites like psychiatry versus primary care in this study. Is the treatment delivered the same? Does that introduce any kind of unmeasured confounders due to treatment at clinical sites? So that's another key design element.

Thank you. Thank you for listening. I've been very lucky, and I've been blessed with some terrific colleagues, including Dr. Edwards, who's part of this meeting today. Over here is actually a snapshot of some graffiti on a sidewalk in Pittsburgh. That's I think a nice message for all of us. This is actually a picture from the inside of the Cathedral of Learning. So thank you very much and have a great rest of the meeting.

Page 101

DR. EDWARDS: Thank you very much, Ajay. That was a terrific presentation. I think we'd all agree that those complex interesting trials probably have a lot of important lessons that we all could take when we're planning future adaptive and pragmatic studies.

I think we just have time for one question or two. There's one that came in. The question is, "How should we weigh patient preferences and patient choice versus the use of randomization to ensure a particular study question is answered? For example, in the STAR-D trial, would you prefer giving patients the choice of augmentation versus starting a new antidepressant and discovering that the majority choose augmentation, or would you prefer to randomize and finally answer the question of which, if either, produces better outcomes?"

DR. WASAN: Well, I think that in a nutshell is the tension between the explanatory and the pragmatic. If you have more pragmatic bend and you want to understand not just what's effective but also what's acceptable to the patients and try to

Page 102

get some insight into what they prefer, randomizing the choice and understanding why patients choose what they choose would be important, but that would dilute your ability to really sort out the, quote/unquote, "efficacy" aspects of each of the treatments.

But that dilution, if you have an adaptive trial in which, say, treatment failures get randomized to something else, that dilution effect of patient choice I think becomes less and less important there. So again, this is about translation -- it's more generalizability. Translation might be the wrong word, but that's what I would think about it.

DR. EDWARDS: Absolutely.

One more question for you, Ajay, comes in from John Farrar, I believe. "Terrific talk." And he wonders, "Given that it's unlikely that any pain treatment works in everyone, is there any way to build in consideration of subgroups that might have a better response and a way of finding them in large pragmatic trials?"

Page 103

DR. WASAN: Yes. That's a great question. One way to do it, if you're thinking about a large pragmatic trial, one of your objectives might be to find which treatments work for targeted subgroups. That gets at it. For instance, you may do a trial in which you do some like phenotyping, and then the patients are randomized based on their phenotype to one of many options.

Let's say you're doing a back pain trial, and the patients who have back pain, plus widespread pain, or back pain plus depression, naturally you would think they could do better on duloxetine versus other treatments because that might be a preferred phenotype just looking at the data. So you might selectively randomize them more to the duloxetine versus the other. That might get at that issue of how do you use phenotyping in order to get at the most pragmatic question, which is which subgroups are going to do the best with which treatments?

Another example is say you have epidural steroid injection, you might selectively randomize

Page 104

people to epidurals if they had back pain plus radicular pain because that's the situation where epidurals tend to work the best. So I think there are some ways to do it, but no one's really done that before, and it's a big opportunity to advance the field that way.

DR. EDWARDS: Terrific. Thanks, Ajay.

I'm going to squeeze in one more question, maybe a brief one. This one comes in from Penney Cowan, and she wonders, "With increasing treatment effectiveness creating reduced pain and increased function, would not that also lend itself to increases in pain flares as patients in the trials become more active?"

DR. WASAN: Yes, yes. Thanks, Penney; always great to hear from you, and that's another thing I think that's underappreciated. Oftentimes, when we're just looking at pain outcomes, we see pain doesn't change, but that's because people self-titrate. They start doing more and they're more active, so they reach the steady-state level that they made. So a 7 out of 10 may be good for

Page 105

that person, and they can do a lot more, and it shows in their function.

My point on the pain flares is that I think, in general, we need to pay more attention to that, like your question alludes to. We need to understand that better and understand the impact of the pain flares and how we model the pain flares analysis. Do we treat the pain flares during the trial? What's this interplay between the change in pain and change in function? I think that's all really important to think about ahead of time.

DR. EDWARDS: Great. Thanks very much, Ajay. Thanks for a terrific talk. Thanks very much to our question submitters as well.

I think we'll now move on to the next speaker on our list. It's really a great pleasure to introduce Dr. Bob Kerns, who will be our next speaker. He's one of the world's preeminent pain psychologists. And not only has Bob himself had a tremendous influence on the field, he's also trained an almost unbelievable number of today's leaders in biopsychosocial pain research.

Page 106

Bob is a professor of psychiatry, neurology, and psychology at Yale School of Medicine. He's a co-director of the NIH, DoD, and VA Pain Management Collaboratory Coordinating Center, which brings together a double-digit number of exciting, mostly multisite trials of nonpharmacologic pain treatments in VA, DoD, and other healthcare systems. Bob and his group at Yale have published hundreds of high-impact papers, covering areas such as comprehensive pain care for veterans, nonpharmacologic pain management, predictors of outcomes of pain treatment, and personalized or precision pain medicine.

So it will be a great pleasure to hear from Bob, who will be talking to us about lessons learned From the NIH, DoD, and VA Pain Management Collaboratory.

Presentation - Robert Kerns

DR. KERNS: Hello, everyone. My name is Bob Kerns. I'm speaking to you from New Haven, Connecticut at Yale University. I'm virtually here with my colleague Cynthia Brandt and Peter Peduzzi.

Page 107

The three of us are multiple PIs or program directors for the United States National Institutes of Health, Department of Defense, Department of Veterans Affairs Pain Management Collaboratory Coordinating Center.

The focus of my presentation today is to inform you about this important tri-government agency partnership, focusing on the conduct of pragmatic clinical trials of nonpharmacologic approaches to management of pain and co-occurring conditions. Our collaboratory is informed by the gap between evidence and practice.

There's growing evidence to support the integrated, coordinated multimodal and interdisciplinary model of pain care that provides for patient activation and pain self-management, and in particular, the evidence supporting the efficacy, if not the effectiveness, of nonpharmacologic approaches and integrative models of care. However, as you all are aware, there's limited uptake of these approaches in routine care in the United States and there are significant

Page 108

barriers to timely and equitable access to these approaches.

The veteran and military health systems supported by the Department of Defense and Department of Veterans Affairs are ideally positioned to address this gap for many reasons, including the fact that they are large, integrated healthcare systems. They are conceptually described as learning healthcare systems that support clinical research integrated into clinical care settings. They have experience with large, multisite clinical trials, and very importantly in the context of pragmatic trials, they have integrated electronic health records that provide data that can be used in the context of the conduct of these trials.

The Collaboratory in fact involves an approximately $81 million investment over six years of funding. The Collaboratory is currently in the beginning of the fourth year of funding for some projects or a third year of funding for some others. You see here the long list of supporting

Page 109

NIH institutes and centers and programs, as well as two branches of the Department of Defense Medicine Research Program, as well as the VA Health Services Research and Development Service.

The objective of our collaboratory is the conduct of pragmatic clinical trials to evaluate whether nonpharmacologic approaches for management of pain and multimorbidities are effective when they are delivered in these healthcare agencies.

Why pragmatic studies? To promote the generalizability of the results and protect rigor and answer questions that inform VHA, the Veterans Health Administration, and the Defense Health Administration about what services to make available to patients with pain throughout their systems. It's also important to note that although there's a focus on the VA and DoD, we hope that the results will inform other healthcare systems about nonpharmacologic approaches and integrative models of care for pain management.

The nonpharmacologic approaches that we're focused on really are a large number that were

Page 110

relatively recently identified by a VA-sponsored health services research and development state-of-the-art conference on nonpharmacologic approaches to chronic musculoskeletal pain. You see here a list of some but not all of the approaches that were found to have support through evidence synthesis reviews and evidence synthesis maps of these approaches.

Some of our trials also specifically focus on integrated models of care. Of course, many of you are aware of the VA and Department of Defense stepped care model that was put in place in policy in VA over a decade ago, as well as a more recent and importantly quite innovative and forward-thinking model of care called the Whole Health Program or All Health Initiative that is designed to empower patients in taking direct responsibility for their health, promoting activities, and to reduce health-damaging behaviors.

Pragmatic clinical trials in our collaboratory, of which there are 11, were all

Page 111

funded by either the NIH, DoD, or VA as cooperative agreements supporting their trials and involved a two-year pilot or demonstration phase, followed by a four-year implementation phase. All of the trials have now successfully transitioned to the implementation phase and almost all are actually involved in recruiting for their trials at this present time.

Here is a list of the 11 pragmatic trials. In this slide, you can see the titles of the pragmatic trials, the principal investigators, and the funding source, either NIH, VA, or DoD. I'll just make a few comments about the breadth of this program. There are trials like the first one listed by Fritz and Rhon that examine the stepped-care model for low back pain management in the military health system.

Another one that focuses on a model of care is the Taylor-Zeliadt trial, in the middle of the slide, that focuses on the complementary and integrative health approaches and uptake of these approaches in the context of the Whole Health

Page 112

Initiative. Similarly, the Seal and Becker trial, in the bottom left of the slide, is focused on this approach as well.

Others of the projects are studying more specific nonpharmacologic approaches like Alicia Heapy's project in the middle of the left column, focusing on an innovative, interactive voice response approach for delivering a CBT-like, self-management program virtually, via the phone, and using IVR that's asynchronous. In fact, it doesn't involve any therapy direct therapist contact. The project by Brian Ilfeld, in the bottom of the middle column, is focused on ultrasound-guided percutaneous peripheral nerve stimulation in the perioperative setting.

The focus of our trials, pragmatic clinical trials, has been highly informed by the Pragmatic Explanatory Continuum Indicator Summary, or the PRECIS-2, Domains, which are listed here on this slide. They focus on eligibility with an emphasis on minimal exclusion criteria to promote generalization and the principle of justice, if you

Page 113

will, to encourage inclusion of people who might be potential targets or appropriate participants in the trials and in receipt of these interventions if they prove to be effective.

Recruitment all occurs in the context of routine clinical care. The settings are all in clinical care settings, primary care physical therapy settings or surgery settings, and so forth. Quite importantly I think for the focus here, as an effort to promote pragmatic trials as opposed to more explanatory trials, the trials all focus on patient-centered outcomes: pain; pain reduction; ability to function in daily life; quality of life; medication usage/reduction; and discontinuation, and all measures, from qualitative and other research, that have evidence supporting their importance in the eyes of people with pain.

Here's a slide with a specific example of the Chiropractic Care for Veterans Trial of Goertz and Long and the PRECIS-2 diagram. You'll see in this trial, the trial is highly pragmatic, relatively speaking, on the continuum from

Page 114

explanatory in the middle of the circle to pragmatic on the outer edges of the circle, with a specific focus on pragmatic outcomes that are important to patients, and maybe a little less pragmatic in terms of their recruitment and follow-up approaches. They've embedded other strategies to enhance what could be done in the context of approaching patients in the clinical care setting either for recruitment or for follow-up.

The Collaboratory is supported by a coordinating center that's based here at Yale. I've already emphasized that there are three program directors. Our coordinating center supports the conduct of the pragmatic clinical trials. We've been involved in their further refinement, initiation, and efforts to help them implement their trials. We support a large steering committee that is comprised of all the members of the coordinating center, the principal investigators, the sponsors, and other important entities such as a separate stakeholder advisory

Page 115

board and a patient resource group.

Our collaboratory coordinating center supports work groups, which I'll show you on the next slide, where teams come together led by co-chairs, faculty-level persons with expertise in particular areas of the work groups, and are supported by program managers and supported by the coordinating center that are populated actually by either the PIs or their representatives. Finally, our efforts are to disseminate best research practices informed by our collaboratory efforts within the Veterans Health Administration, Defense Health Administration, and more broadly, in the academic and scholarly scientific literature.

Here's a list of the work groups: biostatistics and study design; phenotypes and outcomes; electronic health record; data sharing; ethical and regulatory issues; stakeholder engagement; and implementation science. I've already told you about the organization of these groups, and ultimately the groups are designed to guide, support, and facilitate further refinement

Page 116

and development of the pragmatic trials and to disseminate generalizable knowledge.

We've had much success and progress in the last several years as we've begun this collaboratory. All projects, as I've mentioned, have transitioned to the implementation phase. All the projects agreed to use the PEG3 as an outcome measure, not necessarily a primary measure but as a secondary measure to promote data harmonization and sharing across trials, with the potential for the whole to be greater than the sum of the parts.

We all agreed on a common definition of opioid use as either an outcome or as a covariate, drawing on data from the electronic health records in the VA or DoD systems. We've agreed to use the AUDIT-C or the PHQ-2 for phenotyping alcohol use and depressive symptom severity, respectively. We've added additional inclusion criteria and phenotyping harmonization as appropriate to individual trials or small clusters of trials.

We've agreed on a common definition of high-impact chronic pain to identify a particular

Page 117

subpopulation of interest, and we've developed standardized approaches to measuring self-reported use of nonpharmacologic and complementary integrative health approaches, which is an important outcome in several of the trials and may be relevant in further phenotyping of the samples in the other trials.

We right off the bat address site overlap in our trials. Literally, these trials cover over 70 sites across the country with over 50 VA-based facilities being involved as sites for the trials, a large number of DoD sites, and a small number of non-VA/non-DoD sites in some of the trials as well. Right off the bat, we realized that many of the trials had overlapping sites in terms of recruitment, so we worked hard to minimize that overlap, to minimize competition for subjects, and to minimize possible contamination across trials.

We had a large push at the beginning around biostatistics advice, and there was a lot of group discussion as well as project-specific consultation. There were specific white papers

Page 118

developed to address missing data and other important biostatistical and study design issues up front. We worked in the ethics and regulatory group to overcome many challenges related to multiple IRBs and ethical challenges in the conduct of these trials and promoted shared learning and identification of best practices. A specific outcome, and great group that's been involved that was formed, focuses on the issue of justice and health equity, and a manuscript has emerged from this really rich discussion that will be submitted soon.

We've addressed issues of stakeholder engagement across the continuum from veterans and service members and their dependents, to senior policy leaders in the VA and DoD, and even outside our organizations. These groups have focused on shared learning and best practices. The discussion was informed by a lot of qualitative data from several of the trials that were collected in the pilot phase.

There's a manuscript that's in press in an

Page 119

upcoming special issue of Pain Medicine, a supplement that actually will highlight the Collaboratory and includes a protocol paper, so to speak, of each of the 11 trials, complemented by several other papers, including this one on stakeholder engagement.

We've tried to promote the active support of leaders at the military treatment facilities and VA facilities through letters from the Collaboratory Coordinating Center to those individuals, and we've built a patient resource group and external board that's actually chaired by retired Lieutenant General Eric Schoomaker, a former army surgeon general who's a great cheerleader for this project.

Data sharing, we've been approved and we hope to be contributing data to the HEAL repository, and we've done a lot of work building out our website. We're quite proud of this, and I hope many of you will take a look at the website that's listed here. Furthermore, we've developed and written responses related to the COVID-19 pandemic on the PMC website and coordinated and

Page 120

supported unified measures to account for COVID-19 impact across the 11 trials, and a paper on this was published in Translational Behavioral Medicine.

We've also crafted written and video responses that address the Black Lives Matter movement and disparities in pain care, leveraging the PMC pragmatic clinical trial PI expertise, particularly Diana Burgess from the University of Minnesota and the Minneapolis VA Health Care System. We've addressed many other issues in the context of COVID-19, and you see them listed here.

I think most importantly, we've built a supportive community that's come together at a time of great challenge and stress, but a great opportunity for learning, especially in the context of this collaboratory and as parts of the VA and DoD as learning healthcare systems. We've coordinated efforts to identify significant changes to the protocols themselves, particularly addressing issues of sampling and recruitment plans, but also changes in the way assessments were to be carried out and the interventions themselves,

Page 121

some of which have shifted from in person to virtual delivery.

There's a manuscript on this, addressing many of these issues that are currently under review. We've developed additional survey questions for participants in the trials regarding the impact of COVID, and we hope to be able to use these data for further phenotyping in our studies.

We've collaborated with our forerunner, the NIH Health Systems Research Collaboratory, to share in solutions and best practices and develop recommendations that go far beyond our collaboratory or their projects, including, by the way, the PRISM, NIH HEAL Initiative PRISM trials that are supported by that collaboratory. We have been involved in encouraging our PIs to be in active communication with their sponsoring agencies as well as the relevant IRBs and DSMBs.

I'll close with just a few testimonials. You can read them here on the slide. The bottom line is they all really emphasize the value of participation in this collaboratory and the

Page 122

opportunities for building, trusting, and sharing relationships and camaraderie to support each other as well as the success of the individual trials.

With that, I want to thank everybody for your interest and participation. Please check out our collaboratory and follow us on Twitter. I'd also point to the upcoming issue of the Pain Medicine supplement, so stay tuned for that. That's due to be published in December or January, and also an existing paper that was published last year in the Journal of Pain Medicine that describes our collaboratory in more detail than I had time today. Thank you.

DR. EDWARDS: Thanks very much, Bob. That was a wonderful talk, and it is a really exciting collaboratory network that you've set up.

I just want to announce to the group that we are running a little bit behind, so we will be shortening the break. It will be just a 5-minute break, so hopefully those of you on the West Coast who are counting on eating lunch can eat very quickly.

Page 123

We do have just a couple of minutes, Bob, and I have a couple of questions that came in, so I'll read the first one here. "The PMC appears to have harmonized on a number of phenotype and secondary outcome variables. Do all the 11 trials also use the same primary outcome; and if not, was that considered?"

DR. KERNS: Thanks, Rob. And by the way, I can't go without thanking you for that overly generous, maybe from my point of view, a little comical introduction, so thank you.

In fact, no. There was a lot of discussion right up front. All of the trials of course were independently peer-reviewed and have DSMBs that are also engaged in helping make determinations about the design of the trials and such issues as primary outcome measures. None of them changed, I think, their primary outcomes as a function of whence they were involved in the Collaboratory through the shared discussions.

What we were able to do, though, was to harmonize around the use of the PEG3. I think

Page 124

maybe one or two of the trials had already proposed to use that as a primary outcome or they used the Pain Inventory as a primary outcome or secondary outcome from which the PEG can be derived. The simple answer is no. Yes, that was considered, but there were many reasons that that wasn't feasible.

DR. EDWARDS: Great. Thanks very much, Bob.

A final brief question, I think, before the break, "Given that all of the collaboratory trials seem to have progressed to the implementation phase, could you comment on what you'd say are the one or two most important lessons that were learned from the planning phases of these trials?"

DR. KERNS: Sure. I think several of the projects actually had plans for qualitative work that involved key stakeholders/patients, but also clinicians in the trenches, which they were going to rely on to actually deliver the interventions, recruit subjects, et cetera, and other policy makers.

I think it was through that pilot work that actually several projects made important changes to

Page 125

accommodate the interests of those key stakeholders, and of course that continues to be important moving forward, as projects actually have needed to adjust their protocols, their recruitment plans and assessments, and even the interventions themselves in the context of COVID and the rapid closing of in-person care and the shift to virtual care, and now the gradual reopening in some parts of the country of that care.

So I would say that the workaround stakeholder engagement and their input was critically important for many of the trials, and I'd highlight that as a key part of what we did.

DR. EDWARDS: Terrific. With apologies to those watching the clock, I'm going to have to squeeze in one more brief question. Another of the world's preeminent pain psychologist trained by Bob Kerns as it happens, Jennifer Haythornthwaite wonders, "Since these trials require massive collaborations, could you comment on the key processes that you and your team used to build the level of engagement and investment that you've

Page 126

achieved?"

DR. KERNS: Yes. I think we got great advice from the Health Systems Research Collaboratory leaders and work group chairs right off the bat, which was the issue of trust was emphasized over and over and over again. I think that even in the context of a pilot phase, there was some inherent competition so to speak. They all were striving to be successful, and their priorities or incentives were about their own particular trials. So we had to work pretty hard, I'd say over the first couple of years, to build this sense of community and sharing, and trust in that context loomed large.

So I think it was through the work groups, and that people started to get to know each other and started to feel more comfortable sharing their warts, their problems. At some point, we were able to actually integrate a plan where every month in our monthly steering committee calls, either three or four projects, PIs, provide updates to the whole community. We added a, quote, "barrier scorecard"

Page 127

where people had to lay out where they thought some of the key challenges for their trials were, with an emphasis on what they were trying to do to address them.

I think that focus on helping people share best practices and lessons learned in a kind of prideful way, albeit with acknowledging the problems in the background, I think really was a good strategy to help build that sense of community.

There's lots more I could say about that, and some of the quotes I think on the last slide speak to that. But I'm glad you picked up on that, Jennifer, because I do think it's not an easy thing with a large community to build that sense of camaraderie, and I think we've been successful in doing that, and I'm happy to talk about that later in more detail.

DR. EDWARDS: Thanks very much, Bob; important lessons for all of us.

Thanks so much to the speakers so far, to the panel, and to those of you in the audience. I

Page 128

think we will now have time for a 5-minute break before we come back with our final speaker, and then time for discussion. Thanks all.

(Whereupon, a recess was taken.)

DR. EDWARDS: Welcome back, everyone. I hope everyone managed to get a short break in.

After extensive research, I've concluded that I now have the pleasure of introducing our only speaker who has an 18-letter-long last name, so we'll hope this goes well.

David Hohenschurz-Schmidt is an osteopath and a neuroscientist, which is a really neat combination of backgrounds. He's a doctoral researcher at Imperial College in London, in the UK. He and his group are doing exciting work on the autonomic nervous system and pain, among other areas, and probably like the rest of you, I'm excited to hear his talk about pragmatic trials of pain therapies, a systematic review of methods.

Take it away. Thanks, David.

Presentation - David Hohenschurz-Schmidt

DR. HOHENSCHURZ-SCHMIDT: Hello, everyone,

Page 129

and thank you to the previous speakers for their presentations so far. I'll be presenting to you the results of a systematic review of trial methodology, which was commissioned by IMMPACT in order to inform this meeting. I've got no financial interest to declare, apart from the fact that I received an honorarium from IMMPACT for this work, and I'm funded by the Alan and Sheila Diamond Trust for my PhD work, which this falls part of.

So rather than speaking of pragmatic trials, I like the formulation of taking a pragmatic attitude to trial design because that tells us that there's not either a very explanatory RCT or pragmatic trials, but rather a continuum of how easily translatable trial results are into real-world decision-making.

Looking at trials where authors said that they had adopted such pragmatic attitude to trial design, we asked how common those self-declared pragmatic trials are in the pain field, which interventions and patient populations they examined, and what the methods employed are, both

Page 130

in terms of general trial design and those relevant to generalizability into real-world decision-making. Ultimately, the aim of this work was to inform the present meeting with a clear and systematically sourced picture of what is currently going on in this field.

Based on 57 studies that met our eligibility criteria, all published within the last couple of years, we can say that pragmatic trials in pain not only study pharmacological therapies, but a wide range of complex pain interventions and management programs, including cognitive behavioral approaches, physical and manual therapy, and surgery.

They're usually designed as comparative effectiveness trials, but the size and center number varies, averaging at around 300 participants. We also identified a range of areas where reporting, especially the reporting which would be relevant to judge generalizability, was deficient, as well as some areas where design and conduct of pragmatic trials was apparently more

Page 131

difficult to achieve than in others. For example, patient recruitment, pragmatic outcome assessment, or follow-ups were often less akin to what you would expect in clinical practice.

Before going into those results in more detail, I'd like to make you familiar with the methodology employed for our systematic review. We followed a preregistered protocol. We really wanted to capture what is going on in this field across the spectrum of pain therapies. To achieve this, we conducted a systematic review, meaning that we went to great lengths to capture all relevant studies.

Our comprehensive search strategy was applied to seven databases and the search was designed to capture any RCT on patients reporting clinical pain irrespective of the pain diagnosis or the therapy under investigation. The only requirements we had was that some primary outcome measure relevant to the treatment or management of people in pain was to be affected, and the trial had to be declared by the authors to be either

Page 132

pragmatic, practical, or part of comparative effectiveness research. That already shows you one of our limitations up front. We had to rely on author self-report.

We also looked at the last couple of years only because we wanted to get a picture of the current practice, not a historical development. We also excluded smaller and feasibility trials, as we were interested in the challenges of running full-scale clinical trials.

Screening and data extraction was then done in duplicate, and importantly we weren't interested in outcome data, but in trial methodology only, so we didn't extract outcome data. Our data extraction focused on a number of fields: key aspects of trial design, pragmatic trial design, and conduct and analysis.

We also assessed how trialists handled the tension between internal and external validity by looking at methods deemed to affect internal validity, such as randomization procedures, allocation concealment, and blinding of

Page 133

participants. Thirdly, we also looked at the introduction and discussion section to see how authors justified and contextualized both the choice of their trial methods, as well as the results that they obtained.

As part of the data extraction process, we rated each trial on what's called the PRECIS instrument that has originally been developed to influence or to facilitate the design of pragmatic trials but has also been used retrospectively to assess how pragmatic a trial is across nine prespecified domains.

As you can see here, each of those domains is rated on a scale from 1 to 5, where 1 signifies a very highly controlled explanatory approach and 5 signifies that this method was very similar to what you would see in normal, everyday practice. The PRECIS instrument has been shown in this retrospective rating to have good inter-rater reliability and is reasonably well able to distinguish an explanatory and pragmatic trial.

As you can see here, the domains range from

Page 134

how patients were selected and recruited, to how similar the setting and the internal organization of the setting and the treatment delivery, to clinical practice, all the way to the choice of outcome measures, the extent of follow-up assessments, and then the primary analysis. We go into that in much more detail later when we talk about the results.

Data synthesis, then, was mainly descriptive, talking about what we found. We also assessed if certain trial methods were more prevalent under certain circumstances. Like I said, no formal risk of bias assessment was conducted because we weren't interested in outcome data, but we did have a look at baseline age data, assessing if there had been a potential problem with randomization and checking if there was heterogeneity between groups in this feature.

The only deviation from our preregistered protocol was the addition of two subgroup analyses. Here, we investigated whether those PRECIS ratings differed between pharmacological and

Page 135

nonpharmacological trials, as well as between chronic and acute pain interventions. We screened around 770 articles and ended up including 57 trials. We may have excluded trials which were pragmatic but too small or not declared as pragmatic by the authors.

Before we look at the individual methods, it may be worth looking at the objectives of those trials. What I did for this was to look at the aims and objectives statement in each trial report and just feed that into a simple word cloud algorithm, so the larger the word, the more frequently used the word was. Unsurprisingly here, the most commonly used words were around concepts of comparing pain interventions with usual care or to assess as comparative or real-world effectiveness.

I'll now be describing the sample and some general trial methods before we then discuss aspects that are more relevant to the pragmatic trial design itself. Amongst those 57 trials, we had 21 percent looking at pharmacological therapies

Page 136

and then a big bunch of studies, as well, looking at cognitive behavioral approaches, surgery, acupuncture, manual therapy, and a few more exotic therapy interventions.

In terms of the diagnosis, as you would expect, the bulk of trials examined musculoskeletal pain, 9 percent looked at pain after medical interventions, and 5 percent at neuropathic pain and headaches. Just one more point here, interestingly, only one study looked at a diffuse chronic pain condition, in this instance fibromyalgia.

We looked at the duration of the pain. In half the trials, chronic pain was studied, but it wasn't possible to even get an idea of how long patients were suffering from pain in about 28 percent of the trials. So that's our first instance of poor reporting, and then we had something else. I'm going to go through this presentation. The medium number of participants was 234 at the point of randomization, with the largest trial recruiting 1,700 participants. We

Page 137

kept the sample size at a minimum of 40 participants per arm.

Almost half the trials -- you can see that on this slide -- were conducted in primary care settings, but we also had around 30 percent each in secondary and in highly specialized tertiary care. Nine percent of the trials took place in community settings, which included patients' homes. As you can see from this slide here, most trials, 79 percent, took place across multiple centers with a median of 5, but a huge range from 2 to 100 clinical trial centers.

Interestingly, only two reports assessed differences between trial centers, and these authors also discussed how those differences may have impacted trial outcomes. I'd argue that's something important when you have a multicenter trial, especially if you have a hundred different settings or even five very different trials, that that should be considered.

Lastly, most trials were funded exclusively from public sources, three-quarters of all trials.

Page 138

Only 12 percent were fully industry funded. As an interim summary from this section, it was interesting to see that, for example, drug trials were much less common than what you'd expect in the pain field in general, plus also that we had so few industry-funded pragmatic trials.

We're now going to talk about the general trial methods employed starting with the choice of the comparator. Like I said, most were comparative effectiveness trials with around half the studies choosing another active specific intervention as their comparator. We also had placebo-controlled trials. Nine percent were placebo controlled. That's interesting because there's currently debate in the field of pragmatic trials of whether you can call a placebo-controlled trial pragmatic given that in the real clinical environment you wouldn't have placebo interventions, so they do exist.

Blinding was not always reported, but where it was, authors reported that patients were blinded to group allocation in about a quarter of all studies, providers only in 7 percent, and assessor

Page 139

blinding was commonly done in the vast majority of trials. Looking at randomization, we can see that half of the trials were individually randomized with half of those using some kind of blocking mechanism. Stratification by site was the most commonly employed stratification method, and in 10 percent of the trials, the unit of randomization was something other than patients, normally treatment centers or providers, but also towns in one instance.

Ninety percent of the trials were designed as superiority trials, and even though you cannot claim equivalence or comparable effectiveness in a superiority trial that fails to show a significant difference between groups, 9 out of 24 unsuccessful superiority trials did put some inappropriate spin on their results by claiming equivalence or comparative effectiveness. We had four studies designed as noninferiority trials, and despite current FDA recommendations, none of those noninferiority trials included a placebo-controlled group. Out of 15 trials with multiple declared

Page 140

primary outcome measures, only 9 addressed the issue of multiplicity in their analysis.

We're now going to look in more detail at the individual ratings on this PRECIS tool. The overall rating was 3.8 across all domains, and looking at a sample of self-declared cardiovascular trials, which was recently reviewed, we had a very similar average rating. That's beyond the scope of this presentation, but going into the individual domains, there was quite a difference in which domain was higher and lower range. As a reminder here, scores closer to 1 are towards the explanatory end and closer to 5 are towards the pragmatic end of this spectrum.

We're now going to have a look at the individual domains, starting with patient eligibility, which asked to what extent the participants in the trial were similar to those who would receive the intervention if it was part of usual care. As part of this, we also looked at reporting guidelines, mainly the standard CONSORT recommendation but also 2008 CONSORT extension for

Page 141

pragmatic trials. Only 68 percent of the studies reported their eligibility criteria as explicitly framed to show the degree to which they included typical participants or, where applicable, providers, or hospitals.

The main reason for low ratings in this domain was the exclusion of common comorbidities. That was the case in about a quarter of all studies and, similarly, common medications were a reason for non-eligibility in 7 studies. Eligibility criteria for providers, such as the minimum amount of years in practice, were confirmed in 6 trials, but that was very rarely reported; so not reported in three-quarters of all trials, similarly for criteria for trial settings. Authors very rarely justified how they chose who would deliver the intervention and where the trial would be conducted.

The question of recruitment asked how much extra effort is made to recruit participants over and above what would be seen in usual care. This was the lowest average rating across all nine

Page 142

domains. What would be considered pragmatic is convenient sampling, where you just include patients who walk through the door anyway. Instead, we can really see here that recruitment was a challenge in trials, and pragmatic trials in particular.

Despite the fact that 58 percent employed some kind of targeted recruitment strategy, such as identification through records or targeted efforts, despite that fact, 27 percent of all trials missed their recruitment target. The PRECIS domain of setting asked how different the setting of the trial and usual care settings are, and organization means how more elaborate the organization and the care delivery was compared to what you would see in usual care.

Again, there's a CONSORT extension item which asks the authors to report key aspects of the setting, which determine the trial results. If that had been done, that would really enable the reader much more to generalize how the results from this trial apply to their particular setting.

Page 143

However, compliance with this item, this reporting item, was very low with only 37 percent of all trials complying. Similarly, a discussion of possible differences in other settings, where clinical traditions have served as organizations and staffing and resources varies from those of the trial, only happened in a third of all studies.

Aspects of the intervention delivery were standardized in 61 percent of the studies, and out of those, 31 percent employed some kind of fidelity monitoring. On the other hand, the extent to which participants had to adhere to the treatment regimen was very flexible, generally, also demonstrated by the fact that post-randomization exclusion criteria such as minimum compliance or lack of adverse events was only present in 9 percent of the studies.

This PRECIS domain, a follow-up is concerned with the frequency and duration of follow-up appointments, as well as the intensity of the clinical assessments compared to usual care. Based on this, we had an average rating of 3.2, again

Page 144

relatively low, meaning that follow-up was often more elaborate than what would be expected in normal practice.

I'd also like to use this opportunity to talk about trial duration. The average, longest point of follow-up was one year, but that ranged from just a few hours in a trial of acute myocardial infarction to 10 years in a trial comparing partial and complete knee replacements. The average attrition of such follow, the average attrition was 15 percent.

In pragmatic trials, the choice of outcomes should reflect what matters to the patient, and that means choosing direct symptom reports or function-related measures over lab tests and surrogate markers. In our sample of analgesia trials, those were mainly patient-reported outcomes relatively well established potentially, and not least, thanks to the efforts of IMMPACT. Objective outcome measures were mainly used, as secondary outcomes were employed in about half the studies. One thing I'd like to add is that amongst those

Page 145

outcome measures, not a single trial assessed or even discussed risk-benefit analyses to any extent.

The primary analysis, the highest rating here, according to the PRECIS definition, would be given for the employment of a true intention-to-treat principle. That was reported in 84 percent of all trials. However, despite this fact, 21 percent did actually exclude participants who did not provide follow-up data or had missing outcome data. So there's a distinction here between what they called intention to treat and what they actually did.

We're briefly going to look at a selected set of reporting items, so they're all CONSORT items. Like I said before, the general CONSORT recommendations were complied with relatively well, however the CONSORT extension for pragmatic trials, reporting with that was relatively poor, which is a shame when you think about the fact that pragmatic trials are designed to inform clinical practice in a very generalizable way, and as a trial author, you're really in a position to put the reader into

Page 146

your shoes and tell them what the trial setting looked like, what was done, and how that might be applicable to other settings.

We look at individual domains here. First of all, the contextualization of the trial amongst other available treatments and justifying why you're doing this trial was only done in a little bit more than half the trials. Similarly, the choice of outcome measures and follow-up procedures was only justified in half the trials. That may be due to the fact that outcome measures are relatively well established in the pain field.

How and if resources were altered in order to conduct the trial was only reported in half the studies, and looking at the last two columns here, very low, somewhere between 34 and 37 percent reporting compliance with a description of the setting features, which may have determined the trial results, and only 32 percent did actually contextualize the trial findings outside of this trial setting.

In our correlation analyses, we did not find

Page 147

any association between average PRECIS ratings and funding source, the number of trials centers, or the type of therapy or pain descriptor. Similarly, the participant blinding status was not associated with PRECIS ratings, funding source, trial size, or the analysis method. Large trials did, however, have less attrition.

When we looked at our sample, it was obvious that a couple of subgroup analyses were required, so we compared drug and non-drug studies and didn't find any difference in average pragmatism ratings. But looking at the individual domains, we saw that drug trials had less protocol flexibility, which probably doesn't come as a surprise that drugs weren't prescribed in a less flexible way.

Comparing acute and chronic pain trials, as a little caveat here, in both of those instances, the number in those categories differs quite a bit. We had more chronic pain trials and we had more nonpharmacological trials that were needed. By comparing acute and chronic pain trials, we had, overall, higher pragmatism ratings in acute pain

Page 148

trials, and looking at individual domains, we saw that much more recruitment efforts had to be made in chronic pain trials, which potentially drove this correlation.

By way of the summary, I'd just like to point out a few less than a few more surprising findings here. Based on our systematic review, we can say that pragmatic trials in pain research exist and that they are relatively pragmatic when you compare them to a different field. We can say that they comply well with general reporting guidelines, and we've seen musculoskeletal pain and non-drug interventions dominate. Potentially related to that effect, we only had very few trials that were industry funded.

Some of the challenges we saw are likely similar to normal traditional randomized-controlled trials, including recruitment and retention. However and importantly in this field, issues with that may interfere with pragmatic aims of generalizability. Also, generalizability, or rather the reader's ability to judge

Page 149

generalizability, may be frustrated by inappropriate reporting.

A few points that we may pick up on, or may wish to pick up on, in discussion, first of all, we may want to decide if and where we need more pragmatic trials. That could be in pharmacological research, so it could be that we want to value real-world effectiveness over safety and early efficacy evidence maybe in certain pain conditions, such as headaches, neuropathic pain, fibromyalgia, which were underrepresented in our sample.

We may wish to think about promoting the features and the domains seen in the PRECIS domain and include that into recommendations. I think something where the PRECIS tool falls short is its lack of focus on the choice of analysis methods. Like I said, you get a high rating for intention-to-treat analysis. I'd argue that there's other information relevant to real-world decision-making such as cost effectiveness and risk-benefit considerations.

Also, something that would make a huge

Page 150

difference to the field would be more deliberate reporting of pragmatic trial methods and the contextualization of the trial and its findings, basically in line with existing CONSORT extension. I also think, though, that it should be made explicit where concessions had to be made in terms of less pragmatism in trial design due to feasibility reasons, how this was compensated for, or how this may have affected the generalizability of outcomes.

Based on that, I'm looking forward to the discussion now, and I'd like to say a very, very big thank you to my supervisors, Professor Andrew Rice from Imperial College, as well as Professor Bob Dworkin, who's put me in the position to be able to produce this work and present it here. Thank you for that. I'd also like to highlight the contribution of Dr. Annie Bethea Kleykamp, who took me by the hand from the protocol development all the way to the analysis, as well as thank you to everyone who helped with the protocol design, data extraction, and analysis. Thank you.

Page 151

DR. EDWARDS: Great. Thanks so much, David. That was a stimulating and intriguing summary and review of some of the recent findings in the field. It was a nice way to end a terrific series of talks. I have a couple of questions for you that have come in, and then after that, we'll move to a full panel discussion.

The first question is, "It sounds like in your set of 57 recent studies that you reviewed, larger trials actually had less attrition." This person seemed surprised by that. "Any idea what might have contributed to that effect?"

DR. HOHENSCHURZ-SCHMIDT: No, that's really a correlation that we found, which we haven't looked into in more detail afterwards. It may be to do with larger trials simply putting more effort into general trial design, and retention is one of the issues which we thought wasn't addressed in many of the trials. Maybe larger or more elaborate trials paid more attention to that.

DR. EDWARDS: Excellent. Thank you. I've got one more question that's come in. Let me see

Page 152

if I can find it. Yes.

"In your excellent talk, you listed trial methods in your review of pragmatic trials. It looks like about 10 percent of them used placebo or sham comparators, as you nicely noted during the talk. Should we really consider those to be in the category of pragmatic trials?"

DR. HOHENSCHURZ-SCHMIDT: Well, yes. I think there's debate around that at the moment like I highlighted. I think your opinion papers this summer and coming up and also being responded to by the authors from the PRECIS tool, they argued that, yes, you can still gain information there, which is relevant to clinical decision-making over and above efficacy. But that's one of the issues, I think, we need to discuss in the group as well.

DR. EDWARDS: Excellent. It's almost like we planned it. That is a perfect lead-in. Thanks to the short break we took, we're now only a few minutes behind, which is great, and we can move to the full panel discussion. I'll let my co-moderator, Dr. Cherkin, introduce some of the

Page 153

first questions for the panel, and we'll look forward to hearing the responses.

Panel Discussion

DR. CHERKIN: Well, thank you, Rob.

I'd like to start off with a broad question for the panelists.

"Do you believe that pragmatic clinical trials represent a paradigm shift from the double-blind, placebo-controlled mostly drug trials that have been considered the gold standard for research in previous decades? If so, what do you believe needs to be done to change how researchers, funders, and the general public understand the value of pragmatic trials for improving care?"

I am not actually being able to view anybody's -- I guess the plan here is that if you want to answer it, you just speak up.

DR. KERNS: I'll start by just saying what's intuitive to me, which is, no, I don't think it's a paradigm shift. I think that there will be continued need to think about advanced trial designs that are more explanatory in nature and to

Page 154

promote the explanatory trials, just as maybe we further emphasize a focus on the pragmatic end of the continuum and encourage funders and investigators to think about the key questions that they want to ask of the trial and design their studies to best address that, that question.

So I don't think it's either one or the other or replacing an old paradigm with a new one.

DR. CHERKIN: Any other panelists have a comment on that?

DR. HOHENSCHURZ-SCHMIDT: Maybe talking about funding, as you've seen from the review, there's very little industry funding going into this. Do other panel members think promoting that over simple efficacy signals and safety signals, is that something that needs to happen?

DR. EDWARDS: Anyone?

DR. DeBAR: I was going to comment to follow up on Bob's answer, but I can also maybe speak to David's for a moment. I completely agree that it's not an either/or but it's certainly a bold plan for efficacy and pragmatic trials. In fact, the

Page 155

foundation, obviously, for efficacy trials can help us to be able to design trials and have some sense of -- I mean, mechanism remains important for what and how you're formulating your questions, but there are also really important things like usual care and measuring that along the way.

That becomes, in some ways, really the important benchmark in these trials rather than placebo. That's very important when you're looking at mechanisms and you're trying to ferret that out. But, in fact, many of the questions that we're asking are things in which real-world care is moving quite quickly, and the benchmarks we need in order to understand the signal of what we're doing against what's happening in these environments is quite different.

I know, Dan, your role here is as a moderator, but you have also, I think, written quite provocatively about maybe the benefits of looking at placebo as part of the interventions themselves. I don't know whether we'll get to that kind of discussion, but I think it could be

Page 156

important.

DR. CHERKIN: Great. Thank you, Lynn.

Rob, do you want to take the next question?

DR. EDWARDS: Definitely. We had one come in from Ian Gilron for the full panel, maybe starting with Scott Evans if that's okay. So Ian wonders, "What does the evidence show about the handling and impact of missing data on the validity and meaningfulness of results from real-world trials? Which are likely expected to have more missing data than do more controlled trials."

Scott, if you don't mind starting, that would be great, and then anyone can jump in.

DR. EVANS: I think it's a great question and, of course, one of the potential concerns with relying on non-traditional data sources. I think it's going to be very trial-specific, and outcomes-specific, and so forth, but if you get a high prevalence of missing data, significant prevalence -- and it doesn't have to be that high to be impactful in terms of threatening the integrity of your study and the integrity of your

Page 157

results -- we're going to run into issues.
I did have one or two slides where I mentioned a project from CTTI, the Clinical Trials Transformation Initiative, in which they were studying real-world data and its uses. One of the things that came out early in that project was we started critically thinking about when and how can you use, say, electronic health records or claims data for trials.
One clarification that became evident was, well, the devil's in the details. How are you going to use that data, and is it low risk or high risk for either missingness or incompleteness? If it is -- for example, if it's an endpoint for a trial -- then that's a major issue. So that particular group moved from, although it's a very important thing to be thinking about, can we actually use these types of data for measurement of patient outcomes?
Maybe there's a hybrid approach, where there's certain data that is perhaps either more reliable that might be lower risk, and that can be

Page 158

used for other purposes. But for things like endpoints, you can't take too big a risk. You've got to get the results on patients if you're going to be doing trials, and that often needs more formal assessment. But maybe we can use the data for other purposes, whether it's for eligibility and recruitment or even during the conduct of the trial for other things.
So I do think it's a major issue. I think the prevalence is going to depend on what you're measuring and where we're looking for the data, but it is a front-and-center concern.
DR. EDWARDS: Excellent. Thanks, Scott.
Any other panel members that would like to add anything about the handling of missing data in pragmatic trials?
DR. KERNS: I would just say even now in the context of a collaboratory and in the context of COVID, where trials are relying on electronic health record data, the data that are missing aren't necessarily missing at random and that there are potentially some aspects of the missingness

Page 159

that can be understood or explained and others that can't. It gets very complicated.
By the way, it changes over the course of the trial, which adds an additional complexity. A couple of our trials started before the COVID lockdown, and now there are issues about how data has re-emerged in the electronic health record as in-person or other care starts to re-emerge and how that period will be addressed analytically in the approach.
DR. EDWARDS: Excellent. Go ahead, Lynn.
DR. DeBAR: I would just add a couple of things. I think this is a really important topic, and I talked very briefly in my talk about what we did to augment the collection of PROs for pain to what was happening in the [indiscernible] system. I think we could do that systematically and really do it strongly. We may end up with enough data without the kind of reactivity when we rely on point-of-service, collect data, and that it could be really useful.
What I mean when I say that is if somebody's

Page 160

coming in and they're getting a BPI or a PEG when they're coming in with a pain flare, and that data is what we're looking at in an outcome study, we're going to get a real different density in that data with people who have very complicated, severe conditions and folks who may be experiencing some improvements.
On the other hand, if we can push out those assessments -- and we have lots of tools to do that for the personal health record and those kinds of devices, IVR -- and we collect it more routinely, and we know what's happening to people in their day-to-day lives -- I think there's much more data. I don't think it has some of the limitations.
I don't know that we'll ever be able to do the heterogeneity of treatment effect analyses and have the samples that are needed to do that robustly if we don't somehow partner with our colleagues in the clinical front line to be sure that what's used there are psychometrically validated tools and that they're delivered in such a way that we at least minimize some of these

Page 161

compounds.
I don't think it replaces some of the more rigorous sort of research caliber measurements that we do, but maybe we do that with a subset, and we really create crosswalks and understand what we can determine with those PROs. But I hate to have us say there's research grade and there's clinical grade, and we really shy away from what we can do in the delivery system, because I think if we don't do that, then we're always going to end up with this divide between what's happening in clinical practice and what we're doing in research.
DR. EDWARDS: Thanks very much, Lynn.
If it's okay, we'll go on to the next question. One just came in from Matt Bair specifically for Ajay Wasan.
"Ajay, in your excellent presentation, the concept of treatment-resistant depression was raised. Is there a similar concept in pain, so treatment-resistant pain, and could that be a potential moderator of treatment effects in our pain clinical trials?"

Page 162

DR. WASAN: That's a great question. I have not seen that concept, but we've talked about this in many IMMPACT meetings, like isn't the patient phenotyped who doesn't get better with anything? Isn't that a phenotype that somehow we should attempt to model and capture? Hard to say because by the time we see patients in the studies, it's hard to know all the different treatments they've had, between physical therapy and different meds.
Of course, defining treatment-resistant depression, it's a pretty narrow definition with antidepressants, but there's also a similar piece in depression because it's not as if medication is the only treatment that they've sought also. I mean, there are a lot of other nonpharmacological treatments for depression which might be equally as effective; so not just seeing a psychotherapist, but even spiritual treatment, if you have someone who goes to church a lot.
So there is a similar issue in major depression, but we don't really have that concept for chronic pain, but it's something we definitely

Page 163

need to think about.
DR. CHERKIN: Okay. I'll ask the next question.
"Dr. Evans had suggested in his presentation trying to retain the rigor of conventional randomized trials as much as possible in pragmatic trials, including blinding. I want to know from the panelists to what extent is blinding possible or desirable in pragmatic clinical trials of nonpharmacologic treatments for pain?"
DR. WASAN: Well, I'll answer that. I think, again, it's a balance issue. Of course it's very hard to blind both sides, provider and patient, with nonpharmacologic, but you can blind the assessor or you can also blind pieces of it. We're starting a trial now where we have some treatments for pain, and then one of the goals is to try to wean opioids, and the doctor weaning the opioids is shielded from knowing what treatment arm the patients were randomized to. I think a term that gets thrown around is shielding where you can, understanding that it's maybe a little bit of a

Page 164

leaky shield and it's not the same as blinding, but there's some value, I think, to shielding when you can do it.
DR. CHERKIN: So beyond trying to blind the outcomes assessor to treatments, to what extent is it really possible to blind the patients receiving the treatment or the clinician providing the treatments and the assistant?
DR. WASAN: Well, I think that's also in the spirit of pragmatic trials. I think that is another reason why in the collaboratory that Dr. Kerns is leading, it's focused on nonpharmacologic treatments. Nonpharmacologic treatments, for many of those reasons, are really much more suited to perhaps being tested in pragmatic trials. I don't know if maybe that's one of the rationales that, Bob, you guys were thinking when you came up with this. That would be my reaction.
DR. CHERKIN: Bob?
DR. KERNS: Well, I didn't come up with the RFA for the collaboratory. But, yes, you saw in

Page 165

one of my slides that this was really informed by a particular awareness of the gap between the evidence for multiple nonpharmacologic approaches and even growing evidence of the effectiveness of certain models of care, in integrated models of care, and the actual uptake, even in integrated systems like the VHA and DHA.

So in that context, again, it's about the question that's being asked. In this case, there was a particularly strong interest in trying to address that gap, and to do that, it seemed like trials that were more pragmatic, as opposed to explanatory on that continuum, made sense because it really was trying to address largely an organizational and patient care delivery question.

Whether it's about nonpharmacologic approaches versus pharmacologic approaches or other interventions, I would say maybe I'm not so sure that that's the case. I think there are key questions about the effectiveness of pain medications that can be better addressed in a pragmatic context as well. Even though it's

Page 166

nonpharmacologic, one of the trials by Brian Ilfeld that focuses on percutaneous stimulation, peripheral nerve stimulation in the perioperative setting, there's some evidence of the efficacy of that approach, but it certainly is reasonable even there.

So you could think about that as being appropriate for an efficacy trial. Does it work? But it seems entirely appropriate to think about it in the context of a pragmatic design as well, which is the case here.

DR. EDWARDS: A quick follow-up to that excellent discussion about blinding, this one's from Jennifer Haythornthwaite. "What about electronic assessments done privately or at home as an alternative to blinded assessments?"

Any comment from the panelists on that?

DR. WASAN: Yes. I definitely should have mentioned that because that's really become the standard now. That's an important point. Many of us use REDCap and other methods to send the PROs directly to the patient. That particularly also

Page 167

helps that issue.

DR. HOHENSCHURZ-SCHMIDT: Just on that note of blinding, it's definitely an issue in nonpharmacological trials. But when we looked at our sample, there were certain questions where maybe blinding was not that important. One, for example, was a chiropractor trial, where the question was, is it more beneficial to have patients in when they have pain or is maintenance care more appropriate? I think you could argue that that's something which doesn't necessarily need to be blinded.

On another note – and I hope I captured that correctly -- my colleague, Dr. Jan Vollert, as well, has just published a paper. Part of the assumption that we talk about here is that subjective outcomes are more susceptible to placebo effects. That seems to be pretty accepted but is not necessarily always well founded. So his recent paper, for example, showed similar placebo response in rheumatoid arthritis trials in both lab markers and subjective outcomes. That's something to take

Page 168

into consideration, I think.

DR. EDWARDS: Excellent. Thanks, David.

We have several good questions from John Farrar about phenotyping and subgroups. I'll just read one of them here.

"Given the variability in pain phenotype and response, do you think that pragmatic trials might focus on testing pain treatment processes rather than individual treatments?" So process consisting of what treatments to start with, how to progress through treatments, what order to use multidisciplinary approaches in, et cetera. I imagine Lynn and Bob and other members of the panel might have some comment to make about that.

DR. KERNS: I'll start. Yes, I think absolutely. Key questions for primary care providers in the VA are exactly those that John is highlighting: what do I do first; next; the best combination in the context of the field to push for multimodal care? And these are all unanswered questions.

As I pointed out, or tried to point out,

Page 169

several of the trials actually are, I think, studying more processes or models of care. Some are asking in that context very pragmatic questions like the chiropractic trial, which is essentially examining dose of care. I think one of the trials uses a SMART design, where it is sequential testing of approaches contingent on response to the beginning level of the trial.

So I do think that some of these really key questions that at least clinicians and providers are asking the research community really are addressed within the context of a pragmatic trial and focus more on these processes, or steps, or the other aspects I just highlighted.

DR. CHERKIN: Okay. I'll ask the next question.

"The presentations did not make much mention of measuring cost or cost effectiveness of pragmatic trials. Is this not important information to have for those who would be considering adopting an intervention that was found successful in the initial pragmatic trials for

Page 170

application in their own setting?"

DR. KERNS: Several of our trials in collaboratory do include a budget impact analysis rather than a more complete cost effectiveness analysis. Those were secondary aims and our tertiary aims, and are being addressed in a few of the trials. I agree that the opportunity to examine cost, money, in the context of these kinds of trials, is a great opportunity and oftentimes missed in the design of trials.

Of course, there's added cost to doing that. The standards in the field have matured, and it's not just a simple add-on. To do it well, there's a need to make a serious investment in the ability of the trial to really do that well. I think, again, maybe the VA and DoD systems as integrated healthcare systems are particularly tailored, but even there, increasingly of course, in the last decade, in the national push in the VA to paying for care in the community to improve access, this makes even that all the more challenging. So I think that's a tough nut to crack but something

Page 171

very important.

DR. CHERKIN: Any other either disagreements or additions to what Bob said?

DR. WASAN: Yes, I can add something there. I've been part of a couple of large comparative effectiveness trials funded by PCORI, and as many of us know, PCORI has this language that says you cannot officially do cost effectiveness analysis. That was part of, I think, how the Obama administration was able to get PCORI funded; it satisfied congressional demands.

So it's difficult to overtly do that, actually, in some of these large PCORI grants, which have been, in the civilian medicine world, one of the only funding agencies for these kind of studies in the past five years.

Then secondly, even if you are trying to do cost effectiveness analysis, it's very hard to get the data you need. For actual true, healthcare costs, you really can only get that from the insurance company directly. Even where I am, there's a partially integrated healthcare delivery

Page 172

system, and even then, from our own insurance companies, it's very hard to get the actual true cost of care. So those become major barriers even though all of us agree this is important to get.

DR. CHERKIN: Lynn, do you have any comments from the perspective of Kaiser?

DR. DeBAR: Yes. I think -- and Bob spoke to this a little bit -- as opposed to where Ajay is finding some advantages to being systems like Kaiser or VA where we have more integrated care delivery and we can do these kinds of things more seamlessly, I think the other place -- and we do cost analyses, really, in all the more pragmatic trials that we do.

I think the methodology around cost analysis may also inform some of these discussions, and maybe we can come back to it because the variability in cost is really quite, quite pronounced. Ethically, you wouldn't be able to do a trial large enough to really power cost analyses, so there are ways in which you look at the relative benefit for the cost, and you can model those

Page 173

things and really think about who your customer is; in this case, a lot of times the operational leaders.

I had talked about the various trade-offs, some of them quite subjective. I think that cost effectiveness analyses really works in a realm where they're straddling what's important for policy and how you come at this with the best tools available. So maybe we'll have chance for further discussions, so thanks for the question.

DR. KERNS: Dan, if I might build on this topic, I appreciate the focus on budget and money, but I place that in the broader context of implementation. Studies focused -- or integrating, or hybrid designs that integrate implementation methods, that ultimately can benefit the field and stakeholders, if there are positive results and products from the trials, the VA is interested, of course, in promoting their uptake.

So if there's an embeddedness of efforts to address a broader array of implementation, barriers, opportunities, and facilitators in the

Page 174

context of these trials, I think that's important. In the context of the collaboratory, it wasn't part of the original design to have an implementation science work group, but that was actually added after the fact, and I think it was a very important and insightful addition to the Collaboratory, and one that I think will benefit or heighten the impact of the trials moving forward.

DR. CHERKIN: Thanks, Bob.

Rob?

DR. EDWARDS: Thanks, everyone. Incoming question from Bob Dworkin.

"For Bob Kerns and others on the panel, in the Collaboratory and in other trials that aren't blinded, are standard methods built into the studies to reduce patient expectations as much as possible; for example, in the way the consent forms and study staff present the study to patients?"

DR. KERNS: I don't know that I would say that's true across the trials, but I think some trials more than others have attended to that issue, actually quite explicitly. It's not that

Page 175

what they're getting is hidden, but I think there is an effort to make sure that people are aware that -- I guess all the trials at some level have existing evidence of efficacy, if not effectiveness, so there is more an issue of people in the trials are likely to be informed in the consent form that there's no particular benefit to them personally for participating in the trials because these services are potentially already available through routine clinical care.

So that's maybe different in some of our trials and in other trials, but maybe actually consistent with the notion of a pragmatic trial being on the far end of the continuum from explanatory trials to pragmatic trials, where the intervention studies already have some evidence of efficacy, if not even effectiveness.

DR. DeBAR: I would just add a different variant to that. I've mentioned that, at least initially, a lot of pragmatic trials were cluster randomized trials. If you do a cluster randomized trial, your question should be one that really

Page 176

addresses that cluster.

Say, for example -- and this may go back to multimodal and the way that things are rolled out -- that what you're really doing is you're going into a setting and you're shifting things at a, let's say, clinic level about how care is organized. If you do that, there are many instances in which you may not consent individual patients. Because it's being done at that level, these are -- and this kind of gets to some of the safety things that were discussed. Many of the components of pragmatic trials are things that have been well, well tested and that have pretty favorable safety findings, so those are not things you're looking at as much.

But if you're randomizing at a clinic level and you can measure things without needing to go to the individual patient -- so for example, you're looking at some of the implementation outcomes, adherence to treatment for example -- those could be trials where you step around that quite dramatically. So in thinking about the full

Page 177

spectrum of trial designs and questions, I think there are instances where that doesn't have to be a liability.

DR. EDWARDS: Excellent. Thank you.

Bob, while we're still on the subject of this question, a follow-up inquiry related to data harmonization in the Collaboratory. Someone asks, "Is data sharing across these 11 trials being discussed or might there be creation of a pain registry for that collaboratory?"

DR. KERNS: Right. So this is a great opportunity to go back and revisit the question I answered earlier about harmonization on the primary endpoints. In fact, I should have mentioned that several trials are already proposed when they submitted their application, proposed using the BPI, for example, or the numerical rating scale pain intensity measure. So even without any discussion within the collaboratory community, there are subgroups of trials that are harmonized just by chance, I guess, a coincidence on some of those measures.

Page 178

Because by virtue of our implementation science work group, which was added -- it was actually added through an administrative supplement that was funded very early on from the HEAL Initiative, and by virtue of that, we are part of the HEAL collaboratory, broadly speaking. As I mentioned in one of the slides, just briefly, we have been invited to contribute, for the trials, data into the HEAL data repository, and their consent forms accommodate to that when they had a consent form.

So there will be opportunities not only to potentially aggregate data across our trials but potentially across other trials in the Collaboratory. I think particularly relevant may be some of the PRISM trials, but there are others as well. So I think there are great opportunities for that.

Of course, there are major issues around data sharing, partly, I would say because we're talking about two big government agencies and sharing data with another one, and a wide range of

Page 179

issues about data safety, security protections, and so forth. So I think it's a great concept, but we'll have to see how that all shakes out.

The bottom line is the Collaboratory Coordinating Center, I should also be clear, is not a data coordinating center. We actually aren't involved or exempt from human subjects research. So we're really supporting individual trials and the PIs, rather than being in the position to create ourselves some data registry or data management system.

So I think, just to be clear, that's the state we're in right now. I think there are opportunities to summarize for some of the projects to partner with other of the projects that are trying to address similar enough questions with the same or similar kinds of outcomes and to actually share data in aggregate, but we'll have to see downstream how that goes.

DR. CHERKIN: Thank you, Bob.

The next question is directed at Dr. Evans, asking him to comment more about the, quote,

Page 180

"desirability of outcome ranking," unquote. "Is this desirability from whose perspective: patient provider or researcher?"

DR. EVANS: Yes, this is a good question, and it can be constructed from either a caregiver perspective or a patient perspective. The purpose of that particular approach, what we call the DOR, desirability of outcome ranking, was for some of the reasons that I had outlined. We could be missing the boat on certain treatment effects by analyzing each outcome separately in its own bucket and then trying to somehow combine those outcome-specific effects in some way. There can be associations and correlations between these outcomes that may affect your interpretation of a different outcome.

So the turn on words that I had used was what we typically do in trials, is we take the patients in the trial, the trial participants, and we analyze their outcomes, but you can gain insight to things if you flip that around and take the outcomes in the trial and analyze what happens to

Page 181

the patients. If you're thinking about pragmatic questions, that's a natural fit when you're thinking about pragmatic questions.

So we had come up with this notion about desirability of outcome ranking. The first time we had done this, it was created by research clinicians who were knowledgeable in a particular disease area to do that, but of course, the next natural question that came, "Are there ways in which you can bring in the patient perspective?" There are ways in which you can do that. So yes, you can do that. It's perhaps a little beyond being able to describe here, but you can certainly look at it from a patient perspective.

DR. CHERKIN: Okay. Well, thank you, Scott. Rob?

DR. EDWARDS: Excellent. We have a question. I'm going to preface this one by saying that I know that the IMMPACT group is planning a future IMMPACT meeting focused on patient engagement in clinical trials that will bring together advocates, experts, and engagement of

Page 182

other stakeholders in order to make some probably much-needed recommendations about how to involve individuals with pain and other advocates in the design of these trials.

So with that in mind, here's the question for Ajay Wasan and the rest of the panel. "Have people with pain ever taken part in the development and designs of clinical trials? This could be a way to address the expectation of care and the outcomes of what is important to people with pain." This question comes from Penney Cowan. Thanks, Penney.

DR. WASAN: Yes. Definitely, especially in the PCORI trials, that's a big part of it. In fact, PCORI requires, and to their credit at PCORI I think it's really changed how we design clinical trials by including patients. From the very beginning when you do your letter of intent, you actually need to describe how you're engaging patients themselves in the design of the trial.

That's also spilled over to NIH trials. Rob Edwards and I are part of -- it's part of the HEAL

Page 183

Initiative -- something called BACPAC, which is a whole consortium and a series of multiple trials for back pain. In that consortium, too, most of the trials have a patient representative -- at least if not one, at times several -- to do exactly what you suggested. So it is becoming more and more frequent, and even expected in some cases.

DR. KERNS: I mentioned that many of our investigators, PIs, get investigators in the VA Health Care System, and many, maybe all of them actually as I think about it, are in VA health services research and development-sponsored centers of innovation, like ours at VA Connecticut called the PRIME Center.

These centers all now have patient engagement committees. There are some patient groups. I think for the most part, the PIs of the projects who came from those settings, which they are probably half of the trials at least, I think all had engaged the groups of patients even in identifying the key questions to be pitched in response to the RFA, and then certainly have

Page 184

continued to involve them.

I mentioned that during the pilot implementation or pilot demonstration phase, I think several of the projects actually had an explicit subproject or pilot that involved engaging patients, even after the project was funded and the study is designed, but in the context of further refining their approach, engaged patients in helping them make some of those decisions.

I also would be remiss if I didn't remphasize what I mentioned briefly in my presentation, which was that we have a separate collaboratory coordinating center-sponsored patient resource group of very distinguished -- we would all be impressed with the credentials of this group, some of whom are academics, but many are patient advocates or involved in other patient entities -- veteran service organizations and so forth -- that serve as resources for the collaboratory and the trials in particular.

So I think it is a growing interest, as Ajay has already mentioned, at PCORI, but even in NIH,

Page 185

and certainly in VA, it's just a standard now in terms of at least health services, if not broader clinical and rehabilitation research.

DR. CHERKIN: Thank you, Bob.

The next question is, "The presenters have described numerous challenges in conducting high-quality pragmatic clinical trials, even with the substantial financial and technical support that has been provided in the context of the pain management collaboratory. These challenges include not only the usual ones of doing a rigorous study but also the complexity of stakeholders, the importance of relationships and trust, and all of that to get all the pieces to work together so that the trial has a chance of producing credible results.

"So even if these trials are successful, it's not clear that the same interventions will be successful even when implemented in other settings, not just because of the setting being different in terms of affecting directly the outcomes, but just the processes of doing such complex interventions

Page 186

to improve care. So given these challenges, what do you see as the potential for pragmatic clinical trials to have a major impact on health care in the future?" It's sort of a reality test.

DR. WASAN: I can start off a little bit. I appreciate that concern. I just want to remind everyone that one of the reasons there's more interest in pragmatic trials or comparative effectiveness research with a large pragmatic component is because of the problems of purely explanatory trials and RCTS and their difficulty in actually changing pain care and having their findings translate to actual clinical care; so to keep that in mind.

I don't know this literature very well, but the majority of the pragmatic trials in this country have really been in primary care you would think. So you'd get to ask yourself which seems to have been more effective or more influential in changing care or improving care? Was it the pragmatic trial or the RCT? And someone who has a primary care background might be able to comment on

Page 187

that. This is where, actually, if we all had live mics, Matt Bair could speak about this since he's a primary care physician as well.

So that's what I would kind of throw back is, yes, they're messy and they're complicated, pragmatic trials, but clinical care is messy and complicated, and there actually may be more relevance of a pragmatic trial, but we don't really know.

DR. CHERKIN: Thank you. I agree with what you said, but I think this question is, really, just addressing the issue of implementation in other settings of the interventions that have been found in an initial pragmatic trial to be effective. There are so many issues in successful implementation that have to do with trust, et cetera, and not just with the particular intervention. So that's where this was coming from.

DR. WASAN: No, no. I agree. It's a generalizability issue. For instance, I'm part of a large pragmatic trial that took an intervention

Page 188

proven in a pragmatic trial to be effective of physical therapy intervention in the UK and do something very similar, a stratified type of care for physical therapy based on psychosocial risk factors in the United States and much different outcomes. So yes, there's a generalizability question. That's how I would agree with you there.

DR. DeBAR: Dan, I would just add -- and maybe I'm not addressing this fully either. But I think that one of the things that we've learned in these is that you really -- I think I referred to it as vertical integration. You're going to move care whether you're doing QI initiatives or you're doing pragmatic trials, but it's really important that the sponsorship be from the highest level down to the level of medical assistance, people that are assisting in day-to-day care.

I think those are part of the lessons learned about how do we do this work and how do we sustain it. Those are going to be built into the very fabric of the trials. I think that delivery systems are chaotic, things are always shifting, so

Page 189

those lessons learned I would argue could be generalizable in and of themselves.

The other thing I was going to add to what was said earlier, which may not be making the mark, is that one of the advantages of doing trials in multiple healthcare systems is that you've got an invariant core, and then you often have some different tailoring, really, about how those things are implemented, which I would argue helps to increase potentially generalizability, and you can figure out what are the things that you really need to hold with some rigor.

I think we all had these PRECIS-2 designs where we said you can be way out on the [indiscernible] or you can be in a little bit closer. Some of those, when they're aspects of the intervention itself, may be things we want to insist on. So anyway, I do think that those questions can be answered in part with some of the design features of our trials.

DR. CHERKIN: Thank you, Lynn.

DR. KERNS: I'll just mention and remphasize

Page 190

stakeholder, stakeholder, stakeholder, stakeholder engagement. The bottom line of our collaboratory in the promise of success is because there's organizational commitment, embracing of these trials, and shared funding of the trials. So there already is an organizational investment in this kind of research because there is a view, I think an overarching view, that the VA in this case or the DoD should be providing evidence-based care, and that's part of the mission explicitly in these organizations, and buy-in is I think important.

I don't want to minimize for a minute the complexities and challenges because, of course, even in a VA system, it's a large, integrated healthcare system, but it's comprised of literally closer to 2,000 points of access and care. There's a saying in the VA, "If you've seen one VA, you've seen one VA." There also is optimism about some level of uptake and implementation of positive findings and results, but generalizing that to the most rural and least academically oriented settings that are under-resourced and so forth, there are

Page 191

still great challenges to all of this.

DR. CHERKIN: Thanks, Bob.

DR. HOHENSCHURZ-SCHMIDT: I'll just add to that. We're talking about generalizability a lot when we talk about it from the design perspective. I think what came out from our review, and if you think about the end users of the evidence, where they get the evidence from, that's usually scientific papers. They don't have access to a research conference, then can talk to the trial designers.

So if you improve reporting, I think you'd have a huge impact on the uptake. If you're private cap, a manual therapist or something, and you're asking yourself a question about dosage, or spinal manipulation, or something, but you can't really judge how the trial was designed, how they implemented the intervention, what the patients wear [indiscernible] -- simple things, sometimes I think the journal of guidelines of word limits are actually in the way of appropriate reporting. But I think at least some kind of online supplement,

Page 192

where things are described in much more detail than what you usually get out of a scientific paper, that's something which would have a big, big impact.

DR. CHERKIN: Thanks, David.

Rob?

DR. EDWARDS: Thanks, everyone.

David, I think this next question will also start with you, although everyone should feel free to chime in. This is from Bob Dworkin as well.

He notes, "A pragmatic trial could be designed, for example, to test whether yoga is superior to acupuncture in chronic low back pain. But as David noted in his talk, the nonsignificant primary analysis cannot be interpreted to mean that the treatments are comparable. So wouldn't it be more informative to test noninferiority with a usual care group in order to ensure assay sensitivity, rather than test those two treatments head to head?"

DR. HOHENSCHURZ-SCHMIDT: What I can comment on is that that's not a common approach even though

Page 193

it's recommended, yes. You have assay sensitivity, and are you sure that your comparator does what it's supposed to do is something that needs addressing. Also, I think including a usual care group adds another layer of information, which is a pragmatic approach.

DR. CHERKIN: Okay. Anybody else?

(No response.)

DR. CHERKIN: Here is a question, "Thoughts on the trial design of stratified equipoise stratum, which is the largest set of different treatments he or she is willing to be randomized over, then randomize the patient to their equipoise stratum, and then all-causal inferences about pairwise comparisons compared to the efficacy would be based within equipoise stratum randomized contrast."

If I understand this correctly, this is kind of related to expectations or preferences and controlling for that. I am not sure. Anybody want to take that one?

DR. KERNS: I'll just say I don't think

Page 194

that -- as I understand the question -- any of our trials have taken into account patient expectancies or, as you put it, what they're willing to do other than they wouldn't consent to the trial if they know they're being randomized, and they don't want one or the other treatment. But I do think that issues about this makes sense to me, and it's an important consideration, and one that's often only tacitly considered in the design of trials and recruitment approaches, but bears greater attention.

DR. CHERKIN: Okay.

DR. EDWARDS: Excellent. Thank you. I'll jump in if that's okay, Dan. I have another question here. This one I believe is from Ewan McNicol, who is a pharmacist here in the Boston area treating chronic pain patients. This one's for Scott Evans, particularly.

"Scott, in your slides on subgroup analysis, you assign equal weight to efficacy and safety. What happens when those two are not equal; for example, pain relief and constipation versus pain

Page 195

relief and respiratory depression? Similarly, what if the side effect occurs on the continuum; for example, mild versus debilitating nausea?"

DR. EVANS: Those are good points. You can prioritize and weigh different outcomes in different ways. There are even ways to do that on a personalized level. We've had examples of that. I think my broader point about subgroups was that the way we typically identify subgroups of patients is based on either efficacy and a single variable, a pain outcome for example, or avoidance of a particular safety problem, but neither one of those actually identifies the subgroup of patients you really want to treat. The patients you really want to treat are those who experience the efficacy and avoid the safety problem; that is a benefit-risk problem, so thinking about our subgroup evaluations to identify those patients from a benefit-risk point of view rather than a predictive marker for efficacy or a predictive marker to avoid a safety problem.

But you are right. Clearly different

Page 196

outcomes are of different importance; point number one, that their level of importance could certainly be interpreted at a patient level. But there are methods in place that we're working on, and we continue to work on, that try to recognize that.

On the one hand, you can do an analysis that does a transparent, prespecified analysis for a particular weighting system or value system in the way you weight or value different outcomes but could also show you an analysis where either individual patients or clinicians could say, "Well, I have a different perspective," and this outcome is more important for me than some other outcome. How would treatments compare under that scenario? You can show how two treatments contrast as you vary your value system, so we have some ideas about how to do that. Thank you.

DR. CHERKIN: Rob, shall I ask the next question?

DR. EDWARDS: Yep. All yours. Go for it, Dan. Thanks.

DR. CHERKIN: This questioner says, "I

Page 197

commend Dr. DeBar on being able to embed pain-specific PROs in the electronic health record with her pragmatic trial. How was she able to do this? Was there a clinical moment to do this?"

DR. DeBAR: Thank you for the question. I think in some ways we were kind of at the right place at the right time. That trial I was referring to was limited to patients with chronic pain who were on long-term opioid treatment. And at the time, the Kaiser settings in which we are working required that their patients be on an opioid therapy plan, and that required a certain frequency of assessment.

At the time that we entered the field and we were doing the preliminary work in the first UD3 year, they were using the Brief Pain Inventory, and it was interesting because I think many of us, particularly those of us who are psychologists, think of that as a really brief measure, but for primary care providers, that was unduly long in the clinical setting.

When we did some of the formative work, one

Page 198

of the early things we found was that they said, "Well, we only asked you the first four questions to get credit," credit so that they were monitoring their patients, and those were the questions that were related to pain intensity.

What they were able to recognize is that when they were just asking the questions around pain intensity, the ensuing conversation with patients was about how can you reduce my pain. You asked me these questions; you must be able to reduce it, where we were really starting to promote that functionality and functional impairment was maybe where we needed to put an emphasis.

So what we were able to do was to work with Kaiser nationally and make what was embedded into Epic, the PEG. The three items, the version of the briefing inventory, we actually included a fourth because sleep impairment was really important to our PCTs, and that was a tool they were really enthusiastic about using because they said we concretely know how to talk to our patients when we ask four questions, and we can talk to them about

Page 199

these questions about functioning.

So I think what it shows is there's a real interplay to what clinicians may be able to really use in the field what's meaningful to patients and what we might be able to promote. The other tools that we had were that we weren't trying to do all of that in a face-to-face encounter, so we really took advantage of what we could push out and have people do remotely online. As I'd noted earlier, the use of interactive voice response was a means of collecting that because it just would have been to resource-intensive to try to get frequent measurements had we not been able to use those [indiscernible] technologies. Thanks for the question.

DR. EDWARDS: Thanks very much, Lynn.

The next question, it looks like is for the full panel, comes from Nat Katz, who wonders, "What is the evidence that pragmatic trials of treatments for pain have sufficient assay sensitivity to detect analgesic effects compared to a control group?"

Page 200

DR. WASAN: I can start out with that. We all know and love Nat Katz, and we know when Nat asks a question, he generally has an answer himself in mind. In the interest of being provocative, I would counter -- I know a lot of people on this webinar might be a little shock by that, but that's not necessarily the point of a pragmatic trial. I think a pragmatic starts with treatments that have already demonstrated efficacy and already have demonstrated assay sensitivity, and then says how do you compare them to each other in a more real-world clinical context?

So I think it's a bit much to expect pragmatic trials to meet their objectives and also demonstrate all of these explanatory components. We have to assume some stuff is proven for that treatment. We have to assume something and then go forward and plan. So that would be my counter to that.

DR. EDWARDS: That's an extremely effective counter, Ajay. Other thoughts from the panelists?

(No response.)

Page 201

DR. EDWARDS: If not, I have another question here, if that's okay with Dan, if he doesn't mind if I jump in with the next one. Also for the full panel, "Are there some conditions that lend themselves to pragmatic trials more than others? For example, is it easier to do a pragmatic trial in patients with musculoskeletal low back pain or knee osteoarthritis pain than in patients with a neuropathic pain condition?"

DR. KERNS: I'm not sure why that would be. In fact, maybe I'm missing the point. Our trials, at least in our collaboratory and the ones I think that David found in the literature, were, for the most part, relatively large trials that were multisite in nature. Actually, I think there's reasonable expectation -- all the studies are powered to detect the facts. It's not intuitive to me why that would make a difference.

DR. EDWARDS: Fair enough. I've got one more question, and then, Dan, I can flip it back to you. I'm going to try to combine two questions into one. It's a question from Jennifer

Page 202

Haythornthwaite. "Do you think the incentives for patients to enroll in effectiveness trials are different than the incentives for efficacy trials, and how might this affect recruitment and retention?" I'm going to combo that, because they both mention incentives, with a question about how cost might influence patient choice, and in pragmatic studies do patients or should patients pay as they usually do for treatments?

Any comments from the panel on patient incentives and cost to patients in pragmatic trials?

DR. WASAN: Well, we do know in pragmatic trials, cost is a big issue because many times the trial doesn't pay for the treatment. It assumes that insurance would pay for that treatment or charges it to insurance. So in the civilian world that becomes a real problem because of co-pays, and access, and prior authorizations. So that's also a confounder to a pragmatic trial and something that needs to be taken into account, so that is a big factor.

Page 203

For instance, if we know that certain co-pays for things like physical therapy, that the majority of patients who have a significant co-pay, a pragmatic trial might pay for that co-pay, which some can argue, well, that's not too pragmatic. But on the other hand, if patients can't get the treatment of interest, then you have a failed trial to begin with. No matter whatever you're trying to test, they have to at least get the treatment you're interested in testing.

So cost and preference make a big difference. Also, the preference part is important because this gets to this issue where I've advocated for patient choice as an arm, particularly in comparative effectiveness and pragmatic studies, because these are all the factors that are crucial. It wraps in the aspect of expectations. Patient choice wraps in the aspect of cost and why patients choose, and that drives, I think, a lot of the pragmatic outcomes.

So there are some unique concerns. I think it's a very good point, that you need to think

Page 204

about those two issues, the cost and the preference, because it is different for explanatory versus pragmatic trials.

DR. CHERKIN: Okay. I wanted to give the panelists a chance to ask questions of one another if they had any. Are there questions you've been keen on asking but didn't feel that it would be appropriate, since you're all on the panel, to ask one another?

DR. KERNS: I put one in the chatbox earlier for Lynn --

DR. CHERKIN: Go ahead.

DR. KERNS: -- and thought that she might be able to address it. I'm sorry there's noise in the background. I'll be quick.

In the current national discourse around race, and racism, and broader issues of equity, and then bringing that to bear on pragmatic clinical trials, and embedded in clinical settings where there are known disparities in care, access to care, and how care is delivered et cetera, even, by the way, in the VA, what are the risks to pragmatic

Page 205

trials?
Maybe the same as other trials, where those same differences and disparities related to underrepresented or disadvantaged groups are being carried into the trials themselves. Is this important to consider, and are there strategies that you can think about?
I guess this is for you, from your experience, but really for anybody on the panel, or the audience even, to think about are there strategies to try to address that without undermining the pragmatic-ness, if you will, of the trial?
DR. DeBAR: You don't ask easy questions, Bob. I think this is a great question. I will also say that this has been an active conversation in the HEAL multidisciplinary working group this year, where there's really a recognition that sometimes because there are so many logistic barriers and challenges, we're not always doing these trials in the clinical populations and the settings where we really most need that.

Page 206

I don't think I have a perfect answer to this by any means. I had mentioned earlier that I think if we can have trials that really build in, as part of the design, some of the things that many of us have experienced as huge barriers in the work we're doing, we might be part of the way there. One of those things that occurs to me is that we are really poor at engaging patients up front sometimes, particularly for various non-pharmacotherapy treatments where there's a lot more that we're asking from patients as well.
So can we build things where part of what we're testing are those engagement strategies where we allow a much longer lead in, and we look at how long it takes to encourage -- and I talked about encouragement designs -- people to engage in a trial. This is a real challenge, but I think we really need to be assured that our clinical staff involved are representing a diversity of communities and really bring those sensibilities to the table.
Penney Cowan had brought up in one of the

Page 207

earlier questions about stakeholder engagement, and I think we're getting better at this. PCORI has certainly pushed us a long ways along this spectrum, but to be really clear about what are the true barriers involved in treatment, what's acceptable to people, and to do a much better job of really addressing those things up front.
With COVID by necessity, we're pushing a lot of things into telehealth, and there are some wonderful things about that; but to really be thinking carefully about what are the additional barriers that that has for people because access to those kinds of modalities are not uniform, and there are some really interesting community studies that are trying to figure out how do you get phones to people that have much more transient life circumstances.
I feel like my answer's all over the place, but to the extent we can build it into our design and to the extent that we're really engaging stakeholders from a range of communities up front, and all the way through as well -- because we're

Page 208

never going to solve all those things, and we need to adapt as we go. But we have a lot of opportunity and a lot of challenges to do that work.
DR. KERNS: I'll just say that I brought this to the attention of, for example -- just to raise the question, to what extent across for people that do have their finger on the pulse of clinical trials and published clinical trials, like David, I might ask David, are commonly underrepresented groups well represented in their pain clinical trials? And I think the answer is not, that we have the same problems in our trials overall as is true in the healthcare system.
I do feel passionate about this issue. As I'm kind of near the end of my career, this is one that I'm very interested in the public discourse about this. By the way, it's not just race, it's gender, rurality, poverty, and as an older person, older persons, and I do think it's really an important challenge for all of us. It's not unique to pragmatic trials, but it did come up in our

Page 209

discourse because we're embedding them in clinical trials where these disparities exist, and doing nothing doesn't seem to be the right solution.

DR. HOHENSCHURZ-SCHMIDT: I think it's great that you're bringing it up, and I'd just like to add on to that. There are actually a few of those trials. We had one in the inner city area looking at less well-serviced populations. We had a few elderly trials, and we had a few rural trials, which I think comes down to the issue that pragmatic trials ask pragmatic questions, and those are real issues on the ground. So it's great that you're bringing it up.

DR. EDWARDS: Excellent. Thank you, and thanks particularly to Nat Katz for the stimulating and critical question he asked before about assay sensitivity. We've now gotten several additional follow-up questions that have come in. I'm going to read two that essentially ask a similar question in slightly different ways.

"Given Ajay's counter to Nat, do you think it would be possible to design a clinical study for

Page 210

a drug in development that can both meet regulatory requirements for approval and provide pragmatic data to clinicians?"

Similarly -- and this one comes from Bob Dworkin -- "There are no existing pain treatments that have shown consistent efficacy in clinical trials, so a pragmatic trial that doesn't include some kind of control group, whether usual care or placebo, to demonstrate assay sensitivity would seem to have very limited value because it's likely the results will be inconclusive or uninformative, unless one treatment's shown to be superior to another, which is not very likely."

So to sum it up and put my own twist on it, should we be designing at least 3-arm trials for these pragmatic studies, which have active treatment comparator, active treatment, and then a controlled placebo or usual care kind of arm in order to preserve the test of assay sensitivity? I'm happy to hear from the whole panel on this one. We've got somewhere around 7 or 8 minutes left if you'd like.

Page 211

DR. KERNS: Running out of steam.

(Laughter.)

DR. KERNS: Obviously, break wasn't long enough.

DR. EDWARDS: Fair enough. We'll try to keep longer breaks for the session tomorrow, and I think Nat Katz wins the prize for being able to ask questions that are maximally stumping but really fascinating and critically important I think to consider.

Well, I'll follow that up. I'm going to put Ajay on the spot specifically because he's a psychiatrist and I'm a psychologist, so we have an ongoing rivalry. So any chance I get to poke him a little, I'm going to take, even in a public forum like this one.

Ajay, if you're designing your next pragmatic trial, say, comparing psychologically oriented physical therapy to some other nonpharmacologic treatment for chronic low back pain, are you going to include a usual care or a placebo-controlled sort of group because you're

Page 212

worried about being able to demonstrate assay sensitivity of those treatments under study?

DR. WASAN: Well, I think usual care with a placebo, each one of those offers different things. We think about usual care as also a natural history control for what might end up regression to the mean, and placebo, actually I think of it really as an active comparator. I think that neuroscience is strong enough that when you give someone a placebo, you're activating enough endogenous analgesia responses that it's not inert. It actually is an active comparator, and with usual care, you're still getting some treatment also.

So even these things we call controls are pretty messy. I think there's so much scientific importance to understand which treatments seem to get patients better that just doing a head-to-head comparison of two treatments that are presumed -- I'll use "presumed" -- to be efficacious is very reasonable to do without the assay sensitivity because demonstrating this, quote, "assay sensitivity" is a very messy

Page 213

situation in itself and rifed with all kinds of confounders because none of our comparators are as clean as we would like.

So that's what I would say. But please, someone on the panel help me out with a better answer.

DR. KERNS: Well, I'll just say also, I think usual care isn't usual care, isn't usual care, and in a system like the VA, usual care is changing over time. Even that term needs to be better defined, as many of our projects are trying to do in our collaboratory as a whole.

Actually, we're convening one of our annual, this time virtual, steering committee meetings in just a couple weeks, and one of the topics is exactly this: can we better define usual care or standard of care in the context of these trials and apply some standard or encourage at least collection of data, or monitoring, somehow to better understand the evolving landscape of usual care in the context of the single pragmatic trial?

Maybe that's a little off base, but it's not

Page 214

just a simple matter of assay sensitivity I don't think.

DR. DeBAR: I would just agree with both Ajay and Bob on this. One of the trials that we're conducting and just getting underway right now seemed to me really straightforward when we proposed it. It's comparing two telehealth arms of CBT-based treatments, learning from our patients that if we were really going to get at folks in rural and medically underserved areas, we had to make those kinds of modalities available. And we've got a usual care arm as well because we know how much things fluctuate over time.

What we didn't anticipate was that COVID would come along, billing restrictions would lift, the ubiquity now of telehealth modalities, and all kinds of things out there promoting themselves as CBT online and telephonic treatment is part of the landscape. So I think it is really critical that we have that usual care arm. I think even with all the noise and pragmatic trials, we still have the opportunity, I think, to look at interventions

Page 215

where we build in much more treatment integrity than is commonplace in a lot of healthcare systems.

So I'm optimistic that when we look at this, even though I think there will be all kinds of reports of people getting various bits and pieces of what on the surface looks like common therapies, we'll be able to see signal through noise. But that's an example, I think, of the migration, if you will, of services over time in a way that we really want to take into account in these kinds of trials.

Adjournment

DR. EDWARDS: Thanks very much, Lynn. That was an awfully nice summary.

I think we're down to the last minute or two before our hard stop. I think at this point we can thank IMMPACT very much for organizing this meeting. Thanks a million to Dan Cherkin, co-moderator for this session, and thanks, of course, to all of our panelists who provided extremely good input and wonderful talks. Thanks especially to David for his gracious understanding

Page 216

of my pronunciation of his last name. But in addition to that, thanks to all the panelists.

I'm happy to give anyone who wants it the final word, and then I think we can wrap up for the afternoon.

(No response.)

DR. EDWARDS: Terrific.

Jen and Valorie, I think we might be all wrapped up.

(Whereupon, the meeting was adjourned.)

B

C

D

E

F

G

N

P

Q

R

S

T

W