IMMPACT XVIII - Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials

---

June 4, 2015

A Matter of Record
(301) 890-4188

L*4343*L*

INITIATIVE ON METHODS, MEASUREMENT, AND PAIN ASSESSMENT IN CLINICAL TRIALS

IMMPACT XVIII

Ensuring Data Quality in Clinical Trials of Pain Treatments Considerations for Study and Conduct

Thursday, June 4, 2015

Willard Hotel
Washington, DC

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

Welcoming and Introductions

DR. TURK: Please silence your cell phones just like you've been told in every movie theater. Microphones are voice-activated. Please speak directly into the microphone. And one thing that we have seen happened in the previous meetings, as you are speaking into the microphone and you want to comment on somebody else who is behind you or the side of you, and you turn away from the microphone, then and we can't you. So even if you want to talk to somebody that's behind, you do need to talk into the microphone directly.

These are voice-activated, which means that you speak or want to speak, it will directly come on. When six people are in there wanting to come on, no one else will be able to enter until someone else stops speaking. So if you keep saying, why isn't it lighting up, I want to speak, it's because there are five or six people ahead of you in the queue.

So it's not as if you won't get in, it's not

Page 5

your phone's not working -- your phone's not working -- microphone isn't working. It's just that there are other people in the queue ahead of you. As people get off, then you'll be able to get back into the queue at 6. It's a little bit awkward, but I think you'll find it works quite well.

Today's lunch will be held in the Buchanan Room, which I can read. It's located next to the room. Speakers, you're going to find this out, if you can't see the screen, it's really hard to see.

So the luncheon is next to a meeting room. Please note tonight's dinner has been moved to The Nest located on the mezzanine level, so they can find The Nest.

Check-out time on Friday is 12 o'clock noon. Restrooms are located by the board room, which means right outside where we had breakfast. If you go outside the breakfast room, they're right outside there, the important information.

For departures, the airport, train stations, taxis will be available in front of the hotel.

Page 6

Please sign up for the taxis. Valorie and Andrea will have sign-up sheets that you can use. If you need any assistance, please stop by the registration desk. If you haven't picked up your -- haven't signed or if you haven't picked up your tent card with your name on it, please do. They should be on the side over here.

Let's see. Any other housekeeping details for you? Not really.

One other announcement though, if anyone of you are having trouble seeing the slides, it's because we found your glasses last night at the reception.

(Laughter.)

DR. TURK: So if it happens to be you and you're having difficulty, we may have your glasses. Valorie and/or Andrea will have those at the registration desk, so hopefully you will be able to pick those up.

That's the basic details. We encourage you to get involved, get active, talk to your friends, your colleagues, new people. There are some of you

Page 7

who are alumni who have been to IMMPACT meetings in the past. We've learned over time, that alumni come back and want to come back. And some of you that are new people are here. We hope that you'll have an opportunity to meet each other, to talk to each other.

The greatest amount of what we do with IMMPACT and ACTTION is learning to communicate, to speak with each other, not just formally. But what we found we intentionally do is lots of breaks, lots of dinners, lots of opportunities for you to speak with each other, because often what's more important than what goes on during the formal session is what happens when you're talking among yourselves and discussing things. And many of you have noticed that. The alumni have seen that happen.

So what's this particular meeting? Since the 18th when I got started doing this with Dr. Dworkin, he and I both had black hair. Things have changed in 18 years. The title for this particular meeting is Ensuring Data Quality and

Page 8

Clinical Trials of Pain Treatment: Considerations for Study Execution and Conduct. We really are addressing a fundamentally important issue for all of us, for all of you, and for the field in general, how to, in fact, ensure the quality of the data.

We can have the best scientific questions. We can have the best understanding of the anatomy and the physiology, and the biochemistry. If in fact, we don't gather the data appropriately, correctly, accurately and validly, then we can't draw any reasonable conclusions in these studies.

So this is a fundamental, probably should have been the first IMMPACT meeting, or first one or two IMMPACT meetings, that began back in 2002, was the first meeting, I believe.

I want to acknowledge the support from a number of pharmaceutical companies. For those that are new to IMMPACT, what you may not know is that when we have people attending from the pharmaceutical companies, every company that supports us is allowed to have one person here. We

Page 9

encourage, strongly encourage, that the people here from industry not be viewed as marketing people, but rather they're here as scientists who understand the kinds of issues we are talking about and that hopefully they will be talking about things that are meaningful, in general, that's broadly relevant not just to their particular companies. But we do thank these particular companies.

If for some reason the logo of your company isn't there -- there were some things changed right at the last minute -- and the same thing will be true when I show you of who's present. I apologize, but I had to leave Seattle. I'm from Seattle. I left Seattle early yesterday, so things have changed, and I've tried to keep things up to date. But if for some reason I've messed up let me know, and we'll fix it on the slides.

So what IMMPACT is not. For those again, the alumni can immediately turn this off. It's not the International Micronutrient Malnutrition and Prevention and Control Program, in case you were

Page 10

wondering. If you're for that meeting you're in the wrong room. This is not it.

It's not the Interactive Massive Model Proximity and Collision Tester. And these are, by the way, all accurate from the Web; these are real organizations that you can look up, but we're not them.

We're not the IMMigrants Public Action Coalition of Trenton, New Jersey. If you're here for that meeting, again, you're in the wrong place. We're not the International Maine Maritime Potato Action Team.

(Laughter.)

Sometimes it feels that way. And we're also not the Double Impact Tae Kwon Do for those that are in to martial arts, although at times, it may feel to you as if this meeting is like that and —

(Laughter.)

DR. TURK: -- we have to find ways to try to keep the meeting organized, and sometimes it feels like that. But that's not the meeting you're at.

What is IMMPACT? Well, it's the Initiative

Page 11

on Methods, Measurement, and Pain Assessment in Clinical Trials, I-M-M-P-A-C-T. If you go to the Web, make sure you have the double M's in there, and it'll ask you, do you really want to say one M? No, you want both M's so you can see everything about the IMMPACT.org.

It's an international consortium, and by international, I want to especially thank of our colleagues who have come from the other side of the pond that have given generously of their time, not only to be here at the meeting but who traveled a great distance. And then there are those of us who are from Seattle like Mark Jensen and myself who also donated a huge amount of time in having traveled across country.

It's an international consortium of academic research, governmental agencies -- and different agencies are listed there -- industry, consulting, research organizations. And I don't like the word "consumer advocate," but I know from Penney Cowan -- I don't want to say "patient advocate," so I haven't come up what the best term to use; those

Page 12

groups that support people who have pain problems.

Penney, is that okay? Did I say that reasonably well?

We're part of the Analgesic, Anesthetic, and Addiction Clinical Trials, Translation, Innovations, Opportunities and Networks, or ACTTION Public-Private Partnership with the Food and Drug Administration.

Now, you'll notice throughout that there are a lot of acronyms. I want to thank my colleague, Dr. Dworkin, who was awarded a special honorary degree from the Society for Acronyms when he has come up with many of these acronyms, but it's helpful. And the reason -- you may wonder why all these double letters and double TT's, and why is IMMPACT only one M.

Well, if you want to go to Google and find out about IMMPACT, if you type in I-M-P-A-C-T, you will find a huge amount of information irrelevant to what you're looking for. Having the double letters sometimes helps you find who we are.

The same for ACTTION. If you type in

Page 13

A-C-T-I-O-N, you're going to get a whole range of different organizations and meetings, and everything you want to know, but not about us.

So one of the reasons, although there are others as well, for the double letters are to try to help people find us if they want to know about us.

Our mission is to suggest methods for improving the design, execution, and interpretation of clinical trials and treatments for pain. They are all about a better research designs, better studies, so that in fact we can draw better conclusions about the kinds of treatments that we are offering to patients, which is the ultimate end-user of anything that we do.

The whole mission is can we find more effective and efficient ways to make sure that treatments get evaluated, and those that have turned out to be appropriate to be able to get into the hands of the providers as soon as possible.

Who is IMMPACT? Well, over the 18 meetings, we've had 200 different participants. Usually

Page 14

meetings run anywhere from 35 to 50 people. This is one of the larger ones at this particular meeting. Some of you have attended multiple ones, as I have said before. Some of the other people have been at multiple ones.

Academic and related participants from 12 different countries have been attending these meetings over time, and the countries are listed there if you just kind of curious about where they come from. We are short on Asia, so will do the best we can in the future to make sure we have opportunities for our Asian colleagues.

Over 85 different academic institutions have had people who have attended these particular meetings. Participants from government agencies at DOD, DEA, EMA, FDA, NIH, SAMHSA, VA; some of you are all here from some of those organizations as well.

We try to make sure we bring together academics, government people, industry and people who have pain problems have their representation. The idea is to try to have people in the same room

Page 15

who are addressing the same kinds of issues but maybe with some different perspectives, and that's the real purpose of why we're here.

We've had 45 different pharmaceutical companies over the history of the 18 years. Some of which have been multiple ones; some of which are new. And they will come and go depending upon the nature of the meeting or where they are in their development stages on the way.

But as you can see, we've had a lot of people. Obviously, we don't have all those people in the room at one time, different people and different organizations are here for different meetings.

We've also had, as I've mentioned consumer advocacy representatives. Again, I don't like the word "consumer," but okay. And we've had five different organizations involved and 1 and three-quarters are here today because people are sort of transitioning from one to another. So we're happy to have all of you attend in case you're wondering.

Page 16

These are the different organizations that have been involved from the government in case you're wondering about what institutions have been involved. Some of them are participants and some of them are observers. And what I mean by that is that, depending upon the charge from certain governmental agencies, some people just sat and observed and didn't have anything that they wanted to add; they wanted to learn from us. Others have been intricately involved in the discussions.

What do we do? Well, as I mentioned, it's been 18 IMMPACT meetings, and I'm sure you can't read these, especially in the back, but this is just giving you an idea the kinds of topics. These are up through the first 14; the next slide will show you the next few. But these are the kinds of topics we cover, everything from outcome domains to statistical models to ways of improving the design of studies to have to interpret to multiple endpoints, et cetera.

Again, I know you can't read them all, but you can find them easily on our website, which is

Page 17

either IMMPACT.org or you can go to ACTTION -- is it ACTTION.org, Bob? ACTTION.org. Again, the most recent meetings, and as I said, the one that you're here for is this particular one.

So these are the meetings. We've tried to be as transparent about these meetings so that on our website for both IMMPACT and ACTTION, we've included speakers, the topics, background presentations. We ask all speakers who have slide presentations to make those available to us and remove any slides that they are not comfortable with for proprietary reasons, but to put them on the website so that anybody can, in fact, get access to those.

The more recent meetings are being transcribed, so those become available. So if anybody wants to know what we're doing -- those of you that are familiar with IMMPACT and ACTTION know that we try to have every one of these meetings arrive at some type of considerations, discussions, ways to help people improve their studies. But obviously it's not just for the people here, so we

Page 18

try to publish these in mainstream journals, typically in the Pain journals, to try to make sure that information get out as soon as possible.

All of you will be invited to the authors of those papers. What typically happens, if you're wondering, is that some couple of people will draft up an initial version, and we will often come back to speakers and ask them to look at the section or to give us a section from their presentation. We then craft a draft, it gets circulated, you get a chance to put comments on those.

With the number of people in this room, 50-plus people, you can imagine that it takes an awful lot of time. We plead with you if you get this and you are considering being an author, that you turn things around in a reasonably timely fashion. Sometime papers are dragged out for much more time than we want them to, and we get pressured, how come you're not faster? It's because we can only turn things around as quickly as the authors are willing to do that.

Welcome, Dr. Hertz, who just snuck in the

Page 19

back.

So you'll be asked. And for those of you who are from companies, I know that there have been times when they've had to have their legal departments look at anything that's we're doing, and that can take some time. So to the extent that you can expedite that, we appreciate that. The goal is within at least a year, hopefully less, is to try to get these manuscripts out. The idea is to disseminate the information. If it's just us talking to ourselves, that's fine, but we really want to go beyond that.

What does IMMPACT do? In addition to the meetings from ACTTION and IMMPACT that we've been talking about, we commission review papers and conduct scientific studies. So in addition to the presentations and the papers that come out of these meetings, we've also contracted some studies to be conducted on certain things. For example, one of the contracted papers was to do a study which involved patients, and they were patients who had different types of pain problems, to be in focus

Page 20

groups for discussions about the important meaningful outcomes were for them, what bothers them, what's not getting picked up.

So because that came up in one of our meetings and we didn't have enough input and enough knowledge about what those people experiences were, to try to make sure that we included them, so we contracted a study in that.

We've also contracted some background papers on the pediatric aspects of pain because one of the things we felt was there wasn't sufficient information out there about pediatric pain assessment. So those are some of the things we've tried to do, and there are a number of others.

Again, if you're interested in any of these things, you can ask me, ask Bob, or go to the website and find out more about those. Everything we do we try to get published. The idea is to get the information out there.

Articles have been cited, and we always toot our horns. So last I went to a Google scholar 4,100 times in over 600 scientific journals

Page 21

published in 14 different countries. So somehow or another, people are getting access to these and learning about these things.

The journals are running anywhere from addiction medicine, women's health, to my favorite veterinary medicine. Somehow or another, what we're saying has something important for veterinary medicine, so that's good. And that's the kind of things we do.

If you're just kind of interested over the time, first meeting was in 2002. I think the first publication came out in 2003. And that's just showing you -- this is unique citations, so sometimes more than one article will get published, which is where the 4,000 number comes from. But this is just showing you over time, and that's up through about mid-May-ish the last time I went to look to see where we are.

So you can see it's been going up steadily over time. It doesn't appear to be that it's going down in any way. I want to thank the editors of one of the journals who's sitting here who has

Page 22

published some of our papers, Mark Jensen, and we've helped the IMMPACT factor of his particular journal tremendously.

(Laughter.)

DR. TURK: So therefore he's begging us for our favors but we fight him off.

(Laughter.)

DR. TURK: Here's the website I mentioned, IMMPACT.org, if you want to go. This is the old website that's now been imbedded within the ACTTION website, but you can see the information that's there, who's on the steering committees, the publications, one of the instruments that we supported, the development of the Short Form-McGill Pain Questionnaire 2. There's also the other information.

So if you get interested in any of the things you've heard, or if you just happen to be browsing and see some particular meeting that that topic was interesting, gee, I wish I had been there, you can go to the website and you can actually download the information, the

Page 23

presentations, as much as possible, the background articles, the citations. So you find out as much as you want about us. Again, the idea is to make this as transparent and as available to anyone who wants to.

ACTTION, which I have mentioned to you what it stands for. Notice that there are two additional A's in there, and we were asked by the Food and Drug Administration to expand what was originally analgesic to also include anesthetic and addiction products. So now we try to bring all that information in as much as possible to this public-private partnership.

Mission of ACTTION, it's a public-private partnership with the U.S. Food and Drug Administration to identify, prioritize, sponsor, coordinate, promote innovative activities with a special interest in optimizing clinical trials that will expedite the discovery and development of improved analgesic, anesthetic and addiction treatments for the benefit of the public health. That's what we are here to do. That's what you are

Page 24

here to do. And you're here to help us do that.

ACTTION.org, there it is. If you want to go to their website, remember there are two T's in ACTTION and two M's in IMMPACT.

Here's the people who are here at last I knew, which was about two days ago. If for some reason there's been some shifting or changing and who is coming are not coming, I apologize, but this is the best I can do. I have highlighted in yellow those people who are either speakers or they're moderators, or they're discussants on different projects. So if you see your name there -- and any misspellings, it's totally Bob's fault. I have nothing to do with this. This is what he gave me. So that's who's here.

Now, the best way to know who's here, and we'll do this shortly, is to let you go around and just tell us in 30 seconds or less who you are, but mostly so people get the a name with a face as they are sitting around. You all should have your name tags as well.

What are the objectives? I've told you

Page 25

these are ready, to discuss important considerations and provide suggestions regarding the execution and conduct of clinical trials to improve data quality. That's what this meeting is all about. That's what we're going to be talking about for the next two days.

We want to disseminate these considerations, observations, suggestions, and research agenda by publication of peer-view articles. The end of tomorrow, when we'll know this is a successful meeting, is when we sit back as we say, okay, was there enough discussion, did we get enough interesting input, ideas, that we, in fact, can come up with some suggestions, considerations, recommendations, that can be considered by people doing their clinical trials.

Let me caution with the word "considered." We have no authority to require anybody to do anything. That's all we can do is to put the information out there of our discussions of an informed group of people who are thinking about these issues. What people choose to do with the

Page 26

information we put out there is totally up to them.

So we don't have any regulatory authority, we don't have any control, we can't require them to do anything. All we can do is say, this is our opinion. People who are here from the government agencies who end up being authors, and they can decide to be or not, they are not speaking for the government agency when they endorse one of our papers. What they are basically saying is their personal opinion from their experience is relevant to this particular topic, and they agree sufficiently.

Consensus, by the way, does not mean unanimity; doesn't mean every exact person agrees with every word in here. It means there was a consensus or there was a group discussion that led to an agreement that this was reasonably close to what they feel comfortable signing off on.

As I said, all of you will be invited to be authors. You can have commentary on that. We do our best when the manuscripts come back to take -- you can imagine 50 authors, the comments.

Page 27

We then tried to integrate the comments. It's always interesting when we've done is, when we've have three people all of whom want the sentence written in different ways and how do we make it work for them. So then we go to Mark Jensen and say, you're an editor, you know how to do this, and we asked him for his consultation.

So that's what our objectives are. In order to accomplish this, we sort of have to do some herding of you. And because of the time pressure, you will find that we tend to do a good bit of pushing. It's all Bob Dworkin who does the pushing. I'm this gentle guy who just sits back and lets it happen."

So how do you herd participants? Well, some notes for you in the gentle art of herding IMMPACT participants that we've learned over the last 18 years. Participants don't like to be herded. In fact, you can't really hurt IMMPACT participants, but that doesn't stop us from trying.

Participants prefer to herd themselves, but they're not very good at it, so sometimes you need

Page 28

a little assistance.

Participants understand that they sometimes need to be herded; however that doesn't make it any easier to herd them even though you don't realize it.

Harsh herding has negative consequences, so we don't coerce. We don't try to do any herding negatively. However, Bob does do a little bit of herding, and there see him in action of what he's trying to accomplish.

In case you don't know Bob Dworkin, that is him without his glasses.

Okay. Very simply what I want to do now and quickly -- first of all, let me do this. Any questions, any comments, either about the housekeeping details; about the purpose of the meeting; about anything I've presented in the background about IMMPACT and ACTTION that you feel you'd like to know? And any questions, Bob will be happy to answer them for you.

(No response.)

DR. TURK: No questions? Okay.

Page 29

What I want to do quickly, and I know this is a big room, and you can say this is going to take some time, but I do think -- and Bob and I talked about this, that we felt that it might be useful, just so people will have the name with a face, with who's around them, so when they see them, they can talk to them and they will know a little bit about them.

So Kushang, why don't we start with you? 12.5 seconds you have to tell us who you are and where you're from.

MR. PATEL: Kushang Patel. I'm a research assistant professor at the University of Washington. Do you want other background information?

DR. TURK: I think that's going to be enough.

Oh. Bob wants more. Bob wants more background -- Bob doesn't want more. Okay.

MS. CHEN: My name is Crystal Chen from Biogen, Cambridge, Massachusetts.

DR. TURK: And your background is what?

Page 30

MS. CHEN: I'm a trained physician but a medical director now with Biogen.

DR. TURK: I'm glad you're not an untrained physician. Okay.

DR. CHEN: Thank you.

DR. TURK: Nat?

DR. KATZ: I'm Nathaniel Katz. I'm a neurologist from Boston, and I'm at a company call Analgesic Solutions and Tufts University School of Medicine.

DR. TURK: Amy?

MS. KIRKWOOD: I'm Amy Kirkwood. I'm a statistician from the Cancer Trials Center in the UK, and we are part of the UCL cancer research. I'm going to talk about central statistical monitoring this afternoon.

DR. TURK: And you're originally from Georgia or Alabama?

(Laughter.)

DR. TURK: I heard an accent.

MS. KIRKWOOD: Yeah.

MS. DOYLE: I'm Mittie Doyle. I am a

Page 31

rheumatologist and vice president of clinical research at Flexion Therapeutics, which is a small company outside of Boston.

DR. TURK: Thank you. Mark?

DR. JENSEN: Mark Jensen, University of Washington, clinical psychologist by training.

DR. TURK: Wen?

MS. NEIBLER: Wendy Neibler. I'm a neurologist by training, and I'm with Egalet Corporation, a small pharmaceutical company based outside of Philadelphia, focused on developing abuse-deterrent opioids.

DR. MARKMAN: Good morning. My name is John Markman, and I'm a neurologist from the University of Rochester in Rochester, New York.

DR. VANHOVE: Trudy Vanhove. I'm a VP medical affairs, Jazz Pharmaceuticals.

DR. FREEMAN: Roy Freeman, neurologist, Boston Beth Israel Deaconess Medical Center, Harvard University.

DR. TURK: Mike?

DR. McDERMOTT: I'm Mike McDermott, a

Page 32

biostatistician at the University of Rochester.

DR. EDWARDS: Rob Edwards. I'm a clinical psychologist at Brigham and Women's Hospital in Boston.

DR. VRIJENS: Bernard Vrijens from Belgium. I'm a statistician by training and specialized in medication adherence.

DR. SCHUETTE: This is the statistician row. Paul Schuette, statistician, Office of Biostatistics in the Center for Drug Evaluation and Research, FDA.

DR. TURK: We put Bob Edwards back with you guys as a psychologist just to make sure things were okay.

(Laughter.)

DR. TURK: Now, I'm going to have trouble seeing in the back, so I'm just going to sort of point. I'm sorry if I can't call your names.

DR. KERNS: Bob Kerns, VA. I'm a psychologist, VA Connecticut Health Care System; Yale University.

DR. ROWBOTHAM: Mike Rowbotham, neurologist

Page 33

at UCSF and scientific director at the CPMC Research Institute in San Francisco; part of Sutter Health.

DR. TURK: Jim?

DR. CAMPBELL: Jim Campbell. I'm a neurosurgeon by training from Hopkins, and I represent a company called Centrexion, which is in the field of pain therapeutics.

DR. JACOBS: I'm David Jacobs. I'm a clinician, Daiichi-Sankyo Pharmaceuticals.

DR. TURK: Ajay, you want to go back to you?

DR. WASAN: I'm Ajay Wasan. I'm a pain physician at the University of Pittsburg.

DR. SKLJAREVSKI: And I'm Vladamir Skljarevski, neurologist and neurosurgeon by training, working for Eli Lilly and Company, overseeing late-stage pain trials.

DR. RICE: Andrew Rice. I'm professor of pain research, Imperial College, London.

DR. MALAMUT: Hi. And I'm Rick Malamut, neurologist and therapeutic area of pain at Teva

Page 34

Pharmaceuticals.

DR. MULIA: Good morning. I'm Sohail Mulla. I'm a clinical epidemiologist at McMaster University in Canada.

DR. HERTZ: Sharon Hertz. I'm a neurologist by training, and I am currently the director for the review division that covers analgesics at FDA.

MS. BURKE: Laurie Burke, Lora Group, University of Maryland, School of Pharmacy, and formerly FDA, where I established the SEALD staff.

DR. KOVACS: Sarrit Kovacs, reviewer with the clinical assessment staff at FDA, which was formerly the SEALD Study Endpoints Team.

DR. TURK: You've changed your name; is that correct?

DR. KOVACS: We did.

DR. FIELDS: Ellen Fields, clinical team leader in Sharon's division at the FDA.

DR. CONAGHAN: Philip Conaghan, professor of musculoskeletal at the University of Leeds and a member of the OMERACT executive.

Page 35

DR. HEWITT: I'm David Hewitt. I'm a neurologist by training, and I'm vice president of neuroscience and pain at inVentive, which is a CRO, and vice president of Medical and Scientific Affairs at inVentive.

DR. EVANS: Good morning. Scott Evans, biostatistics, Harvard University.

DR. CARR: Dan Carr, a physician and professor at Tufts University where I direct their program on pain research, education, and policy.

DR. TURK: Lee?

DR. SIMON: Good morning. Lee Simon, a rheumatologist, a member of the OMERACT Exec and some other involvement, and a consultant in clinical drug development.

DR. TURK: Lee, why don't you mention what OMERACT is, just because some people may not know the acronym.

DR. SIMON: So there's something called OMERACT, which stands for Outcomes Measurements in Rheumatology, and it's been in existence since 1992. They have had every-other-year meetings

Page 36

since then.

We have 54 working groups all from below, meaning everybody who wants to work in an outcome measurements system, in any disease state in rheumatology, proposes to us. We're a non-membership organization.

They have certain criteria about how to get this done, the evidence that needs to be done. Every-other-year meetings are consensus meetings leading to Adelphi process for certifying or giving approval to whatever is proposed based on evidence We have 975 publications. I think you may have beat us in number of publications, but we only have one area as opposed to yours.

Our next meeting is in Whistler in May of 2016, and then after that it's in Australia sometime in 2018. We have a group of people from all over the world, and we are delighted to continue to work with IMMPACT/ACTTION in coming up with outcome measures in pain.

DR. TURK: Thank you. The reason I asked Lee to give you that background other than just

Page 37

because of the acronym was at the last meeting that we had for IMMPACT was jointly with the OMERACT group, which was the physical function as outcome measurements and clinical trials. Last year, roughly at this time, we had that meeting.

Yes, Judy?

DR. TOCKARSHEWSKY: Good morning. Tina Tockarshewsky.

DR. TURK: Tina.

MS. TOCKARSHEWSKY: Good morning. I have, the past several years, nearly a decade, been serving as the presidency of the Neuropathy Association. Recent years, I've also been a member of the Interagency Pain Research Coordinating Committee.

My time with the association is winding down in the next couple of weeks, and I'm winding up on my own consultancy of working on strategic communications, such as patient community engagement, advocacy work, continuing in the health care sector. And I've also been working on future articles for industry trade magazines.

Page 38

DR. TURK: Thank you.

DR. COSTELLO: Good morning. I'm Ann Costello. I'm trained as an oral-maxillofacial surgeon. I'm with the FDA Center for Devices and Radiological Health, and I'm the pain expert for our center.

DR. JUGE: I'm Dean Juge. I'm a pharmacist. I'm a regional medical director at Horizon Pharma. Most recently I was involved in patient-reported outcomes research for a company and also associate professor at University of Alabama, Birmingham, in biotechnology.

DR. CHEUNG: Good morning. Raymond Cheung, clinician from Pfizer in New York.

DR. SINGLA: Hi. I'm Neil Singla. I'm an anesthesiologist by training, and I work with Lotus Clinical Research, which is an analgesic research site and CRO.

MS. COWAN: Hi. Penney Cowan, founder and executive director of the American Chronic Pain Association for the last 35 years.

DR. KOPECHY: Good morning. Ernest

Page 39

Kopechy. I'm a pediatric clinical pharmacologist by training, and I'm the head of clinical development at Collegium Pharmaceutical.

DR. DEVINE: Hi. I'm Eric Devine. I'm a clinical psychologist at Boston University where I do clinical trials for addiction.

DR. SESSLER: Morning. Nelson Sessler from Purdue Pharma. I'm a pharmacist in the medical affairs group and recently focused a lot on risk management and pharmacovigilance.

DR. UPMALIS: Good morning. I'm David Upmalis. I'm with Janssen research and development. I'm a physician by training.

DR. ALLEN: I am Rob Allen. I'm a neurologist by training. I do clinical consulting with drug development and currently working with inVentiv Health.

DR. GILRON: Hi. Ian Gilron. I'm a professor of anesthesiology and director of clinical pain research at Queen's University in Kingston, Canada.

DR. FARRAR: Good morning. I'm John

Page 40

Farrar. I'm a neurologist and epidemiologist, clinical epidemiologist, at the University of Pennsylvania, interested in pain and clinical trial design for many years.

DR. WITTER: Good morning. Jim Witter. I'm a rheumatologist and medical officer at the rheumatic diseases section of the National Institutes of Arthritis and Musculoskeletal and Skin Diseases. I'm also the chief science officer for PROMIS.

DR. TURK: Which stands for?

DR. WITTER: Patient Reported Outcome Measurement Information System.

DR. TURK: Thank you.

DR. DWORKIN: Hi. I'm Bob Dworkin at the University of Rochester. And please, please, please don't believe a single thing Dennis says about me.

(Laughter.)

DR. TURK: Bob is one of the nicest guys I've ever met.

(Laughter.)

Page 41

DR. TURK: Totally ethical, intellectual giant in the field.

(Laughter.)

DR. TURK: As you heard as we went around, this is really an impressive group, impressive in a number of different ways. In the range of healthcare disciplines, from dentistry to neurosurgery to pharmacology to neurology, women's health. We've covered I think rheumatology, one of the largest ranges I've heard. We have psychologists, epidemiologists, biostatisticians. We've got multiple companies involved. We've got advocacy representatives.

I think this is really a wonderful group of people. I'm awed at the qualifications and status, the knowledge base that people bring to this. I'm really looking forward tremendously to this particular meeting. I think it's going to be extremely exciting.

We are going to do a little bit of herding, so I apologize for that, but to keep things on target, we'll try to move things along. We'll try

Page 42

to have lots of discussion sessions.

By the end of the day tomorrow, you're going to love Bob. He's going to be really sweet. He's going to really work with you to help you craft the beginnings of this manuscript that we're going to circulate. And now I want to turn this offer to Bob.

For the person who isn't seeing the slides, remember, the glasses have been found and we have those available to you.

Bob?

DR. DWORKIN: Thanks, Dennis. Welcome, all of you. I just want to reiterate Dennis' welcome. It's really a great pleasure to introduce our first speaker, who is Dr. Sharon Hertz. As she mentioned, she recently became director of the FDA's division of Anesthesia, Analgesia and Addiction Products, and that enormously pleased many of in the room and elsewhere when she was appointed the director of the division at the FDA.

The other I have to say about Sharon is to acknowledge that from the very beginning of IMMPACT

Page 43

and ACTTION, she's been a steadfast, stalwart supporter, and Dennis and I have greatly appreciated that. Also her mentorship in this whole –- what is now over 12-year-saga. And so thank you very much, Sharon.

Her first lecture –- her presentation is going to be called, A Regulatory Perspective on Threats to the Integrity of Analgesic Clinical Trial Efficacy Data.

Presentation – Sharon Hertz

DR. HERTZ: Good morning. Can I have the next slide, please?

I could say that –- oh, always the disclaimer. So, here it is. I don't know why you'd want me here if I wasn't with FDA, but these opinions are mine and not those of my agency. On a serious note, I will say, for the record, that there will be no conversations about any particular product or development program. There will be no advice given or sought from me or to me. This will simply be a scientific discussion, and that is my purpose for being here, to participate in that.

Page 44

Next slide, please. So you can see I'm no good at acronyms. I'm going to have to leave that to Bob. But this is just a quote from one of the articles that were a part of the background message.

(Pause.)

DR. HERTZ: It's the article by Colin Baigent, and it just describes the concern regarding safety when there are errors in the design conduct, data collection, or analysis of trial data; potential safety issues for the person in the study but also for the future recipients of the drug product. But beyond that potential -- you know, safety is always first and foremost, but really there are a lot of threats beyond that.

The ability to demonstrate efficacy is really the threat that I've seen manifested mostly commonly when they're having problems with clinical trial integrity, or data integrity, because that's the setting in which my experience has been in terms of catching those problems. And this has, in fact, resulted in substantially increased time to

Page 45

get new products to market.
So it's also, simply put, a waste of resources. It's not easy to get these clinical studies done. We do have a lot of patients with pain in this country, but as you all know, they're not limitless. So there are many reasons why it's important for us to get this right from the start.
So just in terms of describing some of the forms of threats to the integrity of analgesic clinical trial efficacy data, there's of course, inadequate study design. If the study can't produce useful information, then none of the data is useful.
We try very hard to work at catching problems early before clinical trials are started, but resources being what they are, the responsibility really comes from whoever is writing the protocol, as well as, to the extent we can, our ability at the agency to provide input.
Sloppy study conduct is extremely frustrating for us to see. And I know that this first item, the training of clinical trial sites is

Page 46

a big pocket that's being discussed. We've discussed it in this setting. One of our former directors for the Office of Biostatistics, that was like one of his primary concerns, Bob O'Neill, is to get these clinical trials site staff, the investigators, all the way down to everyone participating, well trained so that they know the protocol, they know what to do, they know how to collect the data, the know how to record the data, everything, and to train the patients. And that's another topic that's being discussed here.
Then the kind of stupid protocol violations by the staff and study patients, patients who really do forget things or make mistakes, and same thing with study staff. And then sloppiness as it relates to unverifiable data and poor audit trails, when things are not well recorded and managed.
So while this is frustrating, this downright puts us over the top when we find that there were intentional actions that negatively affect data integrity. These are issues that we see in analgesic studies, but really all of our

Page 47

therapeutic areas in the division, but they're not specific to our therapeutic areas. These are broad concepts.
Deceptive subjects; that's one of the articles, the professional subject. Or the subject who simply really wants to get into the study for whatever reason.
Fraudulent data, and I'm going to give you an example of what we suspect was fraudulent data. Intentional failure to adhere to a protocol. Yeah, I got an example for you. And improper handling of data. I got one of those for you too. And deviation from prespecified analyses. Well, you all are going to smirk because I know how many of you want to do that at the end of the day. But we can fix that one as long as the data has been collected properly and locked properly. That one can always be fixed.
So this is a question. We've never been able to prove it, but I think this is an example of investigator fraud. And I've left out all details about the drug product because it's really

Page 48

irrelevant to the story, and I don't want to put anyone on the spot.
But we had a submission come in with three efficacy studies. They had very similar design; two were successful, one failed. Ha. Slam-dunk, right? You've replicated your finding of efficacy, so there really shouldn't be too much of a question.
We started looking at this, though, and we started noticing some things that were a little funny. First of all, the two positive studies were not in the U.S., and the U.S. study was an abysmal failure to differentiate from placebo.
Well, okay, it happens. But we also started looking -- I know what this drug is. It's not its first in class, so we have a history of how this type of product behaves. And gee, that's a really good effect size. Change from placebo of 30 on a VAS? How many people have actually ever seen that happen except perhaps with a single dose opioid post-op study or something really phenomenal? I mean, 30 points. Wow. So we were

Page 49

really impressed by this, but also a little concerned.

Then we noticed one other thing, which is -- I didn't highlight it, so I don't know if any of you have picked up on it already, but lo and behold, oh, man, very little placebo effect.

Well, this happens to be a clinical setting, which is known for an extremely troubling placebo effect, and we can see that that was potentially part of the problem with the U. S. study. The placebo effect had a pretty good sized change, pain intensity difference was 30 points.

So we decided we needed to look over all what we were seeing. So we had a very large effect size, larger than inspected. Yeah. The successful studies had a higher baseline pain intensity, yet it worked better. All right. Sometimes that's helpful for sensitivity.

But, ha, they had a higher pain sensitivity and didn't use rescue or any of the non-drug treatments that were available. And there was less of a placebo response with no placebo patients; not

Page 50

one having a meaningful change, meaningful onset of pain relief.

Well, it started to sound a little too good to be true. So we checked the study demographics looking for things that could explain it. We looked to see if there was a particular site driving the effect. The two foreign studies had the same clinical sites.

Then we started doing an analysis of were we remembering things correctly, so we started looking at other programs that had similar studies, similar drug products. Yeah, we didn't see anyone else who had a similar placebo response, one that low, nor anything close to similar effect size. We even had prior studies of this drug.

So we always do routine site inspections. We have a couple of algorithms we use to check sites. Some of them are pretty basic. Site enrolls a whole lot of people, we're going to check it out and make sure they were doing things properly. There are some other factors that can go in based on the site if somebody's been -- if we've

Page 51

been concerned that an investigative site has been doing things a little funny in the past, we may go double check that.

This was simply based on high enrollment numbers. And the first site, one of the first things the inspector found was that all of the source data had been transcribed from the primary investigator for legibility, and then destroyed. Well, that's a problem with that. That's no-no. Never do that.

This site enrolled 21 subjects in both studies. The study involved people who had an injury. Well, 14 of these 21 subjects who were in both studies, at least a month apart, were injured on the same day twice and were enrolled on the same day twice. You can see this is a pretty good enrollment from this site, 55 in the first study, 35 in the second.

Then the investigator started seeing common surnames and addresses. There were pairs and triplets who were injured with the same injury, on the same day, in the same home, and were enrolled

Page 52

in the study on the same day. Some of them that happened in both studies.

We just excluded that site from the analysis. I mean, after that first bullet, it really didn't even matter what was going on afterwards. But we already had a lot of questions about this program, so we went and looked for other patterns that were similar, and they were there: same-day enrollment with related subjects, or subjects sharing an address, multiple subjects enrolled with the same issue in both studies.

We asked the applicant, "Didn't this make you wonder?" And they said they spoke with the investigators, and they felt that multiple members of the same family or household could sustain the same injury on the same day, and because people in this country were more active than people in the U.S.

(Laughter.)

DR. HERTZ: You guys are laughing, but we were having -- this was serious at the time. We were about to disrupt a major program that had come

Page 53

in for a marketing application. It just so happens that if this had been U.S. sites, we would have actually tried to verify the existence of these people. We can't do that in this particular country. There are privacy laws that precluded verifying the existence of individual subjects. So we had to just end the study, end the inspection at this point.

So we never proved fraud. And it was suspected that there may be a problem, and that's why there's a question mark; there's no proof. But all of these factors and the comparison to the U.S. site was enough so that this didn't go through. And honestly, I got to believe that somebody in the company knew this was coming. They pushed back as hard as they could; I don't blame them. I was a lot of money spent on this, but we couldn't use that data.

This is another product. This is an unusual situation because there was one study site for two studies. That's probably not a good approach. Well, on inspection we found that there

Page 54

was a failure to record key safety variables because the investigator felt that the protocol asked for too much, so he just didn't do it, even though there was a research assistant present specifically to monitor dosing. So it's not like this with a particular burdensome request. And they didn't have the equipment that was mandated by the protocol to establish a baseline parameters.

Well, once you see this for a safety data, you are also going to start to question the efficacy data. We requested some additional data and found more and more problems as it unfolded, and we just basically -- we couldn't characterize the safety of this product, and it's still not on market.

This is a long, very frustrating story for everyone involved, and this product had routine inspections on the first review cycle, found a number of problems. There were problems with protocol deviation. There was a problem with failing to report those violations in the study report. We found them on inspection. That's a

Page 55

no-no. And then there was also what appeared to be some accidental unblinding at the sites.

So this was not approved for a cycle. So after a lot of discussion, another study was conducted. Again, we routinely inspected two of the sites. This time we also inspected the applicant because there have been some issues with that prior study report.

Lo and behold, statisticians had extracted data to create some SAS data sets. Unfortunately they did that before the study database was locked, and it turned out there was an important variable that was unblinded; it was the treatment assignment.

(Laughter.)

DR. HERTZ: Well, the company caught wind of this, and I think what they probably did what was appropriate. I think they immediately blinded the variable. They got rid of the data sets, not the data, the SAS data sets that had been created, and they interviewed the people involved. And it looked as if, in fact, most of the people who

Page 56

potentially had access didn't really even get to it, get to look at the data, nor did they have interaction with the sites or the critical outcome data.

Unfortunately, when this happened, it was about a year before the NDA came in, and we were never notified, and we found out about it on inspection. And there were no audit trails for how the data was managed after the fact. It was all -- they deleted it, and I get that. You're scared; holy smokes, you're just ruining us. We've been going through this; we had to do a whole another study. I mean, I get all that. But you already have a problem with data integrity from a first cycle of review, you come up with a problem.

As soon as you say, nobody has access to this, the next thing should be, contact FDA. Work with us. What we will do is help you confirm that this lapse doesn't have an impact. But without an audit trail on how this was managed, we couldn't confirm anything. Attestations were taken from the people involved, but none of them were still with

Page 57

the company at the time of the NDA inspections, so we couldn't talk to them. It's not on the market yet.

I could sort of summarize this as saying make sure that data integrity is intact, everything involved in that. But obviously, in the running of the clinical studies large multicenter, multinational studies, things are going to happen. But I think the key for us is to plan how to, to the best way possible, try and put practices into place to limit problems. There is no excuse for sloppiness; that's a planning thing.

If you plan on training proper approaches -- and clearly, there's no excuse for intentional integrity lapses, although as applicants and sponsors, I know it's not always possible to know in advance if investigators may run rogue. But when things happen, the next thing that I can say is give us an opportunity to help you create the support you need to limit the damage.

So we're going to find out about it sooner

Page 58

or later. Maybe there are examples that we never uncovered But then an employee gets pissed off and goes to the LA Times or -- these things come out all the time. So I think the other thing I would say beyond attempting to plan to minimize is once something happens, which over time may be inevitable, let's establish a way to try and salvage what you've got.

That's what I have, so.

(Applause.)

DR. HERTZ: Are we doing questions or should I sit down?

DR. TURK: Why don't we just take a couple of questions and leave most of the questions for -- we've left 45 minutes to an hour to have a panel discussion.

So just a couple of questions for Dr. Hertz at this point.

Yes, John?

DR. FARRAR: Sharon, great examples, and it never ceases to amaze me that the thing we try and teach our kids -- which is being in trouble is

Page 59

bad, but lying about it is worse, it also has eliminated a few presidents along the way -- just never gets learned.

The side effect, though, of some of this -- and I would be interested in your comment because we're interested in trying to facilitate this process. But the side effect is that when you participate in the trial, you get 40,000 queries about a period on a page that shouldn't be there and other things.

There is another side to this, which is there can be an over-control, or an attempt to over-control. And I'm wondering if you have thoughts about how to implement enough control, but not so much as to be onerous. No trial is ever perfect obviously, and I take to heart what you say, which is, if you have problems you need to report them. But I wonder what your thoughts are on trying to sort of balance that issue.

DR. HERTZ: Well, you guys are at an advantage because when I was in practice, I really only had a couple of very small clinical trial

Page 60

experiences as an investigator. So I didn't experience that sort of thing, and I don't know the extent of it.

John, is the intensity of that coming from the sponsor, is it coming from FDA inspections?

DR. FARRAR: No, it's from sponsors.

DR. HERTZ: Well, I think that there is a worry then, so I think that perhaps in an attempt to maintain data integrity, it might be a little bit misfocused. Maybe the effort needs to be more on some of the other issues, planning, training, and a little bit less on that sort of thing.

We have a number of -- we have one guidance on monitoring as a way of maintaining integrity, and then there's the ICH document. Those were included in the backgrounder. I believe there's a draft guidance being considered also on this topic.

But I guess I would -- well, one of part of that may also be the greater use of electronic forms may limit some of that technical stuff, and we're seeing more and more of that. But I would just say that that clearly seems like resources

Page 61

misdirected. But I'd rather have them brow-beat you over knowing the instruments, understanding criteria. I like some of the things that we're seeing in terms of blinding about criteria so that people can even unintentionally sway. Yeah. I don't know what else.

DR. TURK: Rob?

ROB: So Sharon, I think your invitation for help is an honorable one, but I was wondering, do you have any examples where -- I'd be interested in our industry colleagues' comments on this. When things go wrong in a trial or within a company, the first impulse is to call the FDA.

(Laughter.)

ROB: I mean, do you --

DR. HERTZ: You'd be surprised.

ROB: Do you have any examples where a company has come to you with an issue relative to fraud?

DR. HERTZ: Yeah. We have been notified. We've been told when there's been –- when their monitors have found improper behavior at a site,

Page 62

intentional or unintentional, and what their plan was in terms of eliminating the site, we discussed whether the data can be used for efficacy, frequently not. But we've come to an agreement on whether it's even suitable for safety because it kind of needs to be reported in some manner; what their plans are to make up the numbers.

So we have. I mean, if you catch it while something is going and you have to eliminate a site, that's something that's potentially fixable. This case with the data breach, they probably could have survived that if we had an opportunity to say, okay, you've cut access, now here is — we're going to have OSI come in and take a look, and work with you on creating the kind of audit trails so we can confirm that these 11 people who potentially had access are now completely off the project and how you're going to -- I think that could have worked.

But I got to tell you, having had a first cycle full of data integrity issues and having a second cycle where we find out through inspection that there was a major potential breach, had us all

Page 63

reeling. We couldn't believe it.

DR. UPMALIS: This is David. There is another aspect of this I think everybody should be aware of as well, is if a site is bad and has done something bad to you, the chances are that that site is involved in other clinical trials as well. And you're not doing your colleagues in the industry any favors by keeping bad sites out there continuing to do clinical trials when they should be closed.

DR. HERTZ: Yeah, and us knowing about it up front, we can do an inspection, we can seek debarment from studies if necessary, have them put on the debarment list. I mean, that does happen.

DR. TURK: All right. So let's hold the rest of the questions, of which I'm sure of there are many for the panel discussion.

Our next speaker before the panel discussion is Dr. Nathaniel Katz. He's on the faculty at Tufts University School of Medicine. He is CEO of Analgesic Solutions. And he's been a second steadfast stalwart supporter of IMMPACT in action,

Page 64

going back to the very beginning. And he will be talking about clinical trial quality, what is it and what approaches can optimize it.

Presentation – Nathaniel Katz

DR. KATZ: Thanks, Bob and Dennis. Thanks so much for inviting me back again. IMMPACT has been one of my more rewarding professional experiences since the very beginning and I guess it was 2001. Was that the first meeting?

Thank you, Sharon, for waking everybody up with those chilling examples of quality problems. And now that everyone is awake, what I'm going to try to do is take it to the next step and present what I hope to be a broader context of what clinical trial quality actually is and how we can go about systematically trying to achieve it.

So first, I'm going to offer you a proposition based on my own experiences and observations, which is that quality, which should be about the ability of the clinical trial to accomplish its intended scientific purpose, the way it has evolved historically over the last number of

Page 65

decades is that there's become what seems to me to be a dichotomy between what I would call regulatory quality, which is, are you following all the rules that you're supposed to be following, GCP, ICH, what have you.

That's what I would call regulatory quality and there's a whole set of checkboxes that inspectors use to ensure that all those rules are being followed.

Then there's a different thing, which is what I would call scientific quality, which is more about, does the clinical trial, as it was designed, conducted, analyzed, and reported, have the ability to actually answer its scientific question?

Now, the peculiarity about this dichotomy is that one would think, or at least I would think, that the purpose of the rules, of the regulatory quality, would in fact be to achieve scientific qualities. So I think, in theory, these should really be the same, but I'll attempt to illustrate for you that I think there's been a divergence, they've kind of gone down different paths. And I

Page 66

think one of our goals ought to be to try to reconcile these two.

So what I'm going to do today is try to present what I think is a concept of scientific quality, and I'll attempt to inform that concept by bringing in insights from a number of different areas, one of them being fundamental principles of experimentation.

We'll try to clarify some definitions. I'll present some results of what I see to be a growing science of clinical trial design and conduct, which many of us here at IMMPACT and ACTTION have been involved with.

I'll also give you a little bit of a teaser on all the work that's been done over the last century or so in manufacturing quality and control, because I think that we in the world of clinical trial quality control have a lot to learn from insights from manufacturing.

So let's dive right in to the real world now. And this is an extract from a recent quality audit that was performed by very experienced

Page 67

quality auditors, quality audits for 30 or more years.

We're in a position now that we run studies and we also act as a site for studies that we are audited a lot. We also do a lot of audits of other organizations, and so I've got access to a lot of these audits. We audit vendors. There are a million audits being done.

So this was an audit of a clinical trial. And I won't say anything more about who was audited and who did the auditing. And these are only the findings from this audit that were considered to be critical. In other words, they were considered to have a major potential impact on the integrity of the clinical trial.

What I want to do is pick out a few of these examples, and I want us all to ask ourselves the question of, is it important, and what is the relationship between these findings and the ability of the clinical trial to accomplish its scientific objective?

So here is critical finding number 1. There

Page 68

was an inadequate security system in the facility. There wasn't a sign-in log in this particular clinical research facility. So was that important? It's important. You have to have some kind of security system. People could walk in and out with drugs, with papers. So having a security system is important. I don't think there's any doubt that that's an important rule.

Is it closely connected to the ability of the clinical trial to achieve its scientific aim? I don't think so. Maybe is it indirectly related if they're bad on security? Maybe they're bad on other things. Sure. All that is possible. But is it closely related to the ability of that clinical trial to achieve its scientific goal? It's not.

Critical finding number 2, the SOPs in this particular organization were two days out of date. The SOP-on-SOPs had a two-year expiration, and that two-year expiration happened to come two days before the quality inspector showed up. And so what this site should have done is monitored that. There should have been a flurry of signatures so

Page 69

that these SOPs were actually "in effect" at the time and they weren't.

So is that a problem? That's a problem. Your SOPs need to be up to date. SOPs are an important thing. Is it closely connected to the ability of that clinical trial to achieve its scientific aim? I don't think so. I don't think you think so.

We can go on and on. There were handwritten notes found on some SOPs. That violates some rules somewhere. There was one version of the informed consent form missing from the trial master file that was someplace else. They had to go find it.

So these things are all important, and they are reflections of quality. But I would call all these things regulatory quality, and I would suggest to you that it's the same for all these findings. They don't really have a close relationship with the ability of the study to accomplish its scientific aims.

I assure you that there's an army of people like this out there. There's a whole industry of

Page 70

quality control inspectors, and this is what they're doing right now at many different clinical research sites and CROs all around the world. And this is the sort of stuff that they produce, and this is important. This is good. This is not a criticism.

The only point I'm making is that this is inadequate and has really very little relationship to the goal of trying to determine whether a clinical trial can accomplish its scientific aim, which is what I call scientific quality.

So you can't help but be reminded when you think about this of that old parable of the drunk looking for his keys under the street lamp where -- and I tried looking for the oldest appearance of this parable. And the first one that I could find that was popular was from the Mutt and Jeff comic strip in June 1942.

Here, you see the drunk saying, "I'm looking for my quarter that I dropped." And the policeman says, "Did you drop it here?" The drunk says, "No. I dropped it two blocks down the street." "Then

Page 71

why are you looking over here?" "Because the light's better over here." So we've all heard this story.

Actually, I don't have the last panel, which is the policeman and the drunk are now both looking together under the street lamp for the lost quarter.

(Laughter.)

DR. KATZ: And maybe you could say that this is the quality inspector collaborating with the clinical research site to look in the wrong place for signs of clinical trial quality.

So I think, like with all things, once they become rules, when we approach quality these days, we're measuring what's easier to measure, but not necessarily what's relevant or what we're trying to get at.

So you've probably noticed by now that I've used the word "quality" a number of times, and I have kind of very subtly tried to introduce some definitions of what the word quality is. But I think it's time now in the presentation to attempt

Page 72

to presentation a definition of quality.

I would say that a meeting about quality, which is what this is, that doesn't attempt to define quality is kind of a waste of everybody's time. So let's at least try to introduce a definition of quality, even though we may not all agree to it.

The definition of quality that I found most appealing comes from this presentation from Leslie Ball, who used to be the head of the Office of Scientific Investigation, the existence of which was already mentioned by Sharon a few minutes ago. And she gave a presentation to the Clinical Trial Transformation Initiative in October 2010. And you can find this presentation on the internet, and I strongly recommend it. It's a very lucid presentation on this topic.

The definition of quality that she proposed is the ability of a clinical trial now to effectively and efficiently answer the intended question about the benefits and the risks of the medical product, et cetera, et cetera.

Page 73

Now, I'll just point out as an aside that she does incorporate the concept of efficiency into her definition of quality. And I actually think that's very important as well. But for the purpose of the rest of my presentation, I'm not going to talk about efficiency at all as a marker of quality. But the ability of the study to answer its intended question, that's quality.

So then you might ask yourself, "Well, what is the intended question of a clinical trial that is the substance of the concept of quality?" And of course, you can ask a million different questions in a clinical trial. And there are all different kinds of questions that are asked. But for all intents and purposes, I'll take actually the same approach that Karen just took, which is to just focus on the measurement of efficacy.

The intended question of a clinical trial is, what is the magnitude of effect of the treatment compared to control or compared to placebo? This is the output of the study that we're trying to produce in a quality way.

Page 74

So this is a measurement task. A clinical trial is a machine that creates a measure -- that's the output -- of a clinical trial. And there's a whole world of metrology out there, and there are multiple societies that have gotten together to try to better understand in a standardized way the whole concept of measurement.

They use a term called a measurement system. And what I would say using that parlance is that a clinical trial can be considered a measurement system. And the purpose of that measurement system is to measure the magnitude of efficacy of a treatment.

So bringing forward these concepts, what I'm going to offer you now is a slightly revised definition of quality, which is, quality means minimizing sources of error that compromise the accuracy of measurement of treatment effect.

That's my working definition of quality, because that's the product, that's the output of a clinical trial. Anything that interferes with accuracy of measurement of treatment effect in this

Page 75

whole conglomeration that we call a clinical trial with all these different moving parts, those sources of error, are the enemy of quality.

Now, to move away from concepts for a second, what I'm going to do is give you an actual example of an attempt to design quality into a clinical trial paradigm. And I'm going to give you the example of bunionectomy because it's familiar, I think, to everybody in this room. And the perspective that I'll give you on the evolution of the bunionectomy model comes from Paul Desjardins. I don't know if you can see that. Many of you know him. He was one of the developers of the bunionectomy model, along with a number of others.

So what Paul will tell you is -- and this is not actually written anywhere that I'm aware of -- that the first five or six clinical trials of bunionectomy failed to discriminate drug from placebo. Ibuprofen looked exactly the same as placebo in those trials.

So when Paul and his colleagues, at Scirex at the time, tried to figure out why that was

Page 76

happening -- everybody knows ibuprofen works for pain, what's going on here, this is some sort of measurement issue, or to use today's word, some kind of measurement quality issue -- they find that the patients were getting different kinds of surgeries. And when they were being assessed for their pain, some had a tight dressing, some had Shewhart charts, more, and some had a loose dressing, which charts less. Some had their feet elevated, some had their feet hanging. Some were going to physical therapy, some were just coming from physical therapy.

So there were all sorts of factors that impact the output of this study, and namely the pain intensity measures that were not being controlled for. And those were those sources of experimental error: type of surgery, timing of assessments, et cetera, et cetera.

Once they standardized all these things, which in the industry you would think about as standardizing a process, once these things were standardized, boom, all of a sudden ibuprofen was

Page 77

better than placebo, p less than .05.
So did the drug all of a sudden start working? The drug didn't all of a sudden start working. The drug was doing the same thing. The drug was having the same impact on these people's pain that it was having all along. The problem was that they were covariates. There were factors that impact on the output of the clinical trial that were not being controlled. That's the problem.
Now, I also want to mention, in case it's not obvious to you, what kind of problem and quality this is and what kind of problem and quality this is not. This is not a statistical issue. There's no amount of statistics that you can do that's going to fix this problem. This is a problem of experiment, a fundamental problem of experimentation, of experimental design.
This is also not a problem of your outcome measure. You could get the SEALD group. You can get Laurie Burke, and you can have a whole team of people together. You can figure out what's the best outcome measure in the world. That is not

Page 78

going to change this problem one iota.
This is a problem of failure to control experimental error, which is something entirely different, and which we've kind of forgotten about a lot in the work that we do. So if you're going to design an experiment that fulfills the definition of quality that I just articulated earlier, that it's capable of answering the scientific question, the first step towards quality is experimental design. And you can design an experiment, but if it's not conducted the way that you designed, well, you haven't really accomplished anything much, either.
So this gets back to the first -- so to restate what Paul and his colleagues did, they went back to Claude Bernard's Principles of Experimental Medicine from the 1850s, which is the first major work on experimental medicine, and they recognized that an experiment consists of two things. There's an input and there's an output. The input would be in this case the treatment intervention, the ibuprofen or the placebo, and the output is the

Page 79

pain score.
So an experiment consists of an input and an output. And to design a good experiment, you hold everything else constant. That's the fundamental principle of the bioassay. And until we do that, you can have all the statisticians, you can have all the measurement experts, you don't have quality.
So to expand from there, what are the other sources of measurement error in clinical experiments? The lack of attention undermines quality, or to put it a different way, what topics do we need to cover if we're going to produce quality in clinical experiments?
This is by no means a complete list. We've already heard about fraud and data fabrication. I just told you something about the initial design of your experiment, which in my view is by far the most important aspect of this. Of course, bad pain measures can be a problem. Inaccurate reporting of pain by subjects is a problem, all different kinds of study conduct problems, as we've heard,

Page 80

covariates that confound this relationship, data storage problems, data analysis problems. You forgot a semicolon in your SAS code.
These are just a short list of all the different elements of quality that need to be attended to.
So if you're going to try to minimize sources of error, you need to have a way of systematically determining what those sources of error are. You could guess what sources of error might be in clinical trials, and I've just guessed at a few, and other people have guessed at a few. Some of those will be right. Some of those may not be right.
So we need a systematic approach to assessing the validity or the importance of different sources of measurement error in order to focus our attention on what's important and what's not important, and also to know when it's been fixed. How do you do that?
There's a variety of different ways that one can try to determine whether a candidate's source

Page 81

of measurement error is actually a source of measurement error in your clinical trial. There's a couple of different ways to do this. One way is what I would call the candidate variable approach. And you'll all recognize this because your names are on some of these publications.

This is what people in IMMPACT and ACTTION have been doing for the last 15 years, which is to try to take an educated guess as to what might be a factor that could influence the ability of the study to measure its intended outcome, what I'm calling quality for today, but what for 15 years we've been calling assay sensitivity.

So you can suggest a candidate variable. Then you can do some kind of a study. You can do it retrospectively. You can do it prospectively. You can be John Farrar, and you can say, "Well, I wonder whether baseline pain variability might somehow reflect error in our measurement system?" And you can go to a study and figure out whether in fact high versus low baseline pain variability actually does impact your ability of your study to

Page 82

achieve its intended goal.

You can do the same thing with subject enrollment rate, as Neal Singla sitting in the back of the room did with colleagues from Pfizer in a published clinical study.

You can look at placebo response rates, as Bob Dworkin and others have done, to see if that as a characteristic of your measurement system interferes with its ability to accomplish this intended result.

You can look at the ability of subjects to report pain accurately, as Mark Jensen has done and as we have done in a number of different studies.

If you do that, you can just one at a time go down the list -- it's almost like the candidate gene approach in studies. You can go down the list and see which factors are relevant and which factors are not. All these factors are, as it turns out.

I just pulled in a little graphic just for fun to show you one analysis that we did where we were wondering about whether patients who use

Page 83

scales accurately are able to discriminate treatments from placebos better than patients who are not able to use scales accurately.

The problem, of course, is that if you have data from a clinical trial, how do you know whether one of those 800 patients in the clinical trial is using the scales accurately or not. So we decided to play with the idea that, if you ask the patient their pain in two different ways, you ought to get a similar answer.

So in this clinical trial, which is an intervention for osteoarthritis, they were asked their pain using the WOMAC pain subscale, very standard, and also using a patient global assessment and measuring pain intensity. And here were the people who had high pain scores on both. Here were the patients who had low pain scores on both. Those patients seemed okay.

Outside these red lines are patients who either had high pain score on the WOMAC, but that same patient at the same time told you their pain was low on the PGA or vice versa. And so it turns

Page 84

out that if you throw out these patients whose pain scores were discordant, which was almost half of the subjects in this 800-patient clinical trial, you increase your observed effect size of therapy by about 50 percent.

This has been shown in multiple of the ways and multiple other analyses. I'm just throwing it out there as an illustration that quality requires that the human subject, who is the measurement instrument in the clinical trial, be calibrated. Otherwise, you're, like, measuring the pH with an uncalibrated pH meter. And you can't possibly hope to achieve your intended goal of your study, at least in any kind of efficient way. You have to overcome it by enormous sample sizes, which is wasteful and, I would say, also probably unethical.

Now, with all this talk about quality, you think that we were the first people sitting in this room to think about quality. And in fact, in the world of clinical research, if anything, we are very latecomers to the challenge of trying to define and identify quality and figure out

Page 85

mathematical approaches to measuring it and controlling it.

This actually began in the 1920s in the United States in engineering. And the pioneer of quality in the world of manufacturing and industry was this guy here, Walter Shewhart. And anyone who works in virtually any industry, automotive, paper, radio, whatever, the name Shewhart is a household name.

So he worked for Bell Telephone in the 1920s, and he had to figure out how to control the quality of transmission of signals in underground cables, which led up to a huge explosion of interest in this in World War II for obvious reasons.

This is his classic work on this topic, statistical methods from the viewpoint of quality control. And he founded this field of statistical process control, which the use of statistical methods to control the functioning of processes in any kind of a system.

He wasn't an awesome communicator. And his

Page 86

main disciple was this guy here, Edwards Deming, who actually was the editor of this book, which is actually a series of lectures that Shewhart gave. And Deming really became the proponent, the worldwide proponent, of these statistical approaches to quality control. He went to Japan and spent many years there after World War II. And he's credited with this whole Japanese approach to quality, which everybody is very familiar with.

The other interesting twist to this story is that these were both engineering statisticians. Deming studied statistics in the U.K., University College London, with R.A. Fisher and Jerzy Neyman, two of the founders of the modern field of biostatistics.

But the peculiarity is that their learning was one way, because he learned from those founders of biostatistics and then used that knowledge to create this field of statistical process control. But there's virtually zero awareness in the field of biostatistics about the use of statistical process control methods in engineering.

Page 87

So -- and there are biostatisticians in the room -- very hard to find a biostatistician that knows anything about statistical process control or uses it. It's even more ironic, because if you went to a pharmaceutical company and you went to the manufacturing plant, they're probably very familiar with these methods and using them every day. But if you go to the other building, where they are analyzing the clinical trial data, nobody has heard of these methods.

I'm not an expert on these methods, either, although I've used them in a number of projects for about a 10-year period of time. And I just want to introduce you to two fundamental concepts of statistical process control.

One is the notion of a process. You're controlling a process. What is a process? A process in this SPC parlance is a unique combination of tools, materials, methods, and people engaged in producing a measurable output, a measureable output, for example a manufacturing line for machine parts.

Page 88

Does that sound familiar? Does that sound like a clinical trial? It's a clinical trial. And the output from the clinical trial is the measurement of the treatment effect.

The main tool that's used in manufacturing quality control is a control chart. And there's a million different flavors of control charts for all different kinds of purposes. You can measure their performance and decide which one you want to use for a certain application.

A control chart is a graphical display illustrating variation typically over time in the output of a process, cell phone defects, sizes, orientation of a transistor on a circuit board.

Typically, a control chart will show boundaries of statistical control limits, where if something goes beyond a boundary, you can say that that process is out of control, and therefore likely to lead to a defective product. This is every day in the world of engineering.

Here are standard control charts that I actually pulled off of the Wikipedia biography of

Page 89

Shewhart himself. And this just shows some engineering process; it doesn't really matter. This is an upper control limit. And so this is the natural variation in that process. And when that variation exceeds this upper control limit, then the process can be thought of as being out of control. And then someone has to go investigate what's going on and get the process back in control.

This is a Shewhart chart, is what it's called. But this is actually from a clinical study, where this is one site in a clinical trial that's being monitored using this Shewhart chart. The top red line here is eDiary compliance of all the patients in that particular research site.

This gray-ish variable here is mean pain intensity scores of all the patients at that site over time. And here on the bottom is variability of pain intensity, week over week, at that clinical research site. And this is out of 20 or 30 weeks of that clinical trial at that site.

So if you were an engineer, you would look

Page 90

at this and you'd say, "Look, at this particular time point," which happens to be about week 7, "the process went out of control." eDiary compliance now exceeded its lower control limit. Mean pain intensity scores followed a week or two after and exceeded their lower control limit.

Variability of pain scores spiked at that particular research site and got into the zone where, according to John Farrar's paper, you would begin to worry about measurement error of your clinical trial. Then for whatever reason, things went back into control about five weeks later.

Now, if this were an assembly line in a manufacturing plant, what would have happened is that there would be alarms, there'd be red lights flashing. There'd be alarms that would go off. The assembly line would be shut down. You'd have a whole team of engineers descending on the manufacturing line to figure out what went wrong with our process, because if we allowed this to continue, we're going to be producing a defective whatever-we're-producing.

Page 91

In clinical trials, we have more of an approach of keeping your eyes closed, and hoping for the best, and then looking, and doing post-mortems after the study was done to figure out what went wrong. And this is basically what I do for a living.

So this is a couple of principles of manufacturing quality control that I think we could learn from. And we have. And this is almost my last slide, so what I decided to do was hire a software engineer from the automotive industry and have him build a system for doing this in clinical trials, which has now been done.

This is just a few sketches of what the system does. This is a Web-based data visualization interface. And you start out with a map of the world of all your clinical research sites. And if any of those key performance indicators exceed their upper or lower control limits, the dot for that site will turn red. If it gets into a control zone, the dot will turn yellow. And if things are fine, that dot will stay green.

Page 92

You can click on that, and then go through to review each individual research site. And you can see what's going on with all the different variables at that site and see, well, geez, that site turned red. What variable was it that made that site turn red? Was it eDiary compliance? Was it protocol violations? Was it adverse event reporting? Was it variability of pain intensity? Was it any of the key performance indicators that we can monitor?

What went wrong? When did it go wrong? You can also see the key performance indicators here on a study-wide level, the red ones being variables that are out of control and also at a site level. You can also click on a variable and rank all the sites in terms of which sites are most out of control with certain variables.

You can also click on the site and look at individual subjects. So if eDiary compliance is the problem, which subject is it that is the one that threw that site over the edge? And thereby, you can achieve what Sharon was talking about

Page 93

earlier, which is not just letting the river flow by and hoping things work out for the best, but identifying problems early and doing early course corrections within the limits of what's kosher in the clinical trial environment to try to keep things on track and not let the study go over the cliff, and only find out about it afterwards.

So I'm only presenting this as an illustration of how there's a lot known about quality control and there's a whole mathematical foundation for quality control that exists in the world of manufacturing. And all you have to do is realize that a clinical trial is a manufacturing process of some type or can be looked at in that way. And the same mathematical principles that are used in manufacturing can be applied to clinical trials, and I think this is where we are going now.

So in summary, what I've tried to leave you with is a framework of thinking about quality in clinical trials as the following. It's the identification and minimization of sources of error that compromised the accuracy of the output of the

Page 94

trial, which is the measurement of the treatment effect.

Quality control rests on a few premises. It rests on evidence that certain variables, which hopefully are amenable to surveillance, if you're going to do something about them, are relevant to the study output, because some things are relevant and some things are not.

We do know about some of the variables that are relevant to the study output already. And I showed you a list of them earlier, and many of us are continuing to work on figuring out what those are.

From a statistical quality control perspective, we can conceive of a clinical trial as a process with many components, including a measurable output, which is a measurement of the magnitude of treatment efficacy.

I think future work is needed to further define what are the relevant variables to study quality and what's the best method for surveillance and correction. And I think the rest of the

Page 95

presentations at this talk will be to focus on specific variables that we all know have a major impact on study quality: medication adherence, for example, measurement of pain.

I hope that we can now view those talks that are coming up in this meeting as specific examples of variables that affect quality control, but that need to exist in a more general context and approach to quality. And that's what I have to say. I hope you enjoyed it.

(Applause.)

Q&A and Panel Discussion

DR. DWORKIN: Dr. Hertz and Dr. Rowbotham, join us for the panel discussion.

So thank you both very, very much for getting the meeting off to what I felt was a great start. I want to ask the first question as chairman's prerogative.

It seems to me that, speaking just for myself, in thinking about this meeting, I wasn't making a distinction that I think now is relevant and helpful. And I just want to see if I'm on the

Page 96

right track. And that is, I think we're talking about three very different things in this meeting.

One is identifying threats to quality sources of error -- you talked a lot about this, Nat, and also Sharon -- so this kind of identification of threats. Another is how to prevent them. Can we set up systems in advance, in designing a clinical trial that will mitigate, lessen, and prevent all of these sources of error?

Then I guess the third thing, that in my head had been all glommed together and never clearly distinguished was, okay, once bad stuff occurs, of whatever type that we've already discussed, what do we do about it when you discover that 1 out of 50 sites had some fraudulent data? Do you just throw out the data from that site and then just go on, analyzing the rest?

So it's that trichotomy of identification, prevention, and dealing with these things after you haven't prevented them. Is that correct? Is that a reasonable way of thinking about the topic of this meeting?

Page 97

DR. HERTZ: Yeah.
(Laughter.)
DR. DWORKIN: Dennis always accuses me of being long-winded, and I think this was the example of that.
DR. KATZ: A slightly longer answer. Yes, although I would add that there's sort of an evidentiary piece as well because you have to know what it is that you're looking for before you identify it.
DR. HERTZ: Yeah, but I also had sort of a different sense of categories developing. I have a very concrete example of what I've done in terms of threats to integrity means that there's a problem with the conduct, whatever, measurement, management of the study. But Nat's approach to quality -- there's almost a differentiation between the concept of a threat to the integrity of the data and how to ensure quality of the data.
There's a little bit of an overlap, but clearly some divergence in terms of -- and I notice that, actually, in the background reading, when I

Page 98

actually got a chance to sort of look at the whole list of what was put together.
I think the contrast between, for instance, John's paper on baseline variability versus the guidance on central monitoring, so it's how do you improve the quality from the perspective of assay sensitivity, improving assay sensitivity, which is something that we've dealt with a number of times and continues to be a challenge, and then also the concept of these other threats external to the nature of the data itself.
DR. DWORKIN: Michael?
DR. ROWBOTHAM: So I think there's a way of unifying both of the talks, and it really comes from manufacturing systems, like the Toyota Quality Improvement System that came also along after World War II.
Under the old system, products would be manufactured, go down in an assembly line, and then there would be a final quality inspection. And if the product was defective, it'd be thrown out. And if it was okay, it would be sent on for

Page 99

distribution and sales.
The problem when you have low-quality systems is you have a very high defect rate. And so you spend a lot of time and materials manufacturing widgets. And at the end, you end up throwing away half or more of them.
So what Sharon was telling us about was really the hazard of waiting until the end, after the study has been conducted, to go back, and look, and see whether or not the widgets met quality standards. And what you find is that, as you start eliminating subjects, eliminating subject sites, study sites because of defects, at the end, you really just have a basic clinical trial power problem.
You started out with a thousand subjects and, after you eliminate all the bad sites, the subjects that were duplicates or their data wasn't collected properly, and you throw all those out, you end up with a hundred subjects left. So now your study is severely underpowered to answer any kind of clinical question, and you have to throw it

Page 100

out.
So that's the hazard of waiting until the very end to do your inspection.
The revolution that came in with what Nat was talking about was the idea of monitoring quality in every single step so that every worker on an assembly line has the power to pull the cord, stop the assembly line until things are fixed, and then, and only then, does the assembly line start up again.
So no one ever passes a defective widget on to the next step in the manufacturing process. And so you don't really need to do that much at the very end because you really already found the early mistakes and corrected those before you get to the end of the line.
So to do that in clinical trials, I think it's obvious with what Nat was showing us, is that they really have central data monitoring, and continuous data monitoring, so that they can find these kinds of anomalies.
So I think, as we go through these

Page 101

discussions the next two days, think about the hazards of waiting until after the fact versus assessing all the important elements along the way. It provides a kind of overall framework for thinking about these questions.

DR. HERTZ: Because I want to go back to the question or the comment that it's not the initial thought, do we have a problem once we contact FDA. But really, what's the outcome of that? One is, there's a fix and there's an adequate audit. And when it gets to a regulatory authority of any country, it'll be noted, and that'll be fine, or it won't be okay.

So that's the one problem. But the other issue is, as you saw in my example, sometimes there'll be an attempt to fix it or ignore it, and then it's the wait-and-let-the-agency-find-it approach, which I got to tell you, colors the whole application and often leads to a lot of suspicion and a lot of extra inspection.

So I guess if we can get the systems in place to identify, to set the system up properly

Page 102

for the most efficient delivery of goods to identify early and in real-time problems, either plan for or develop fixes in real time.

Then the one thing that I would add is to have that extra link to the regulatory authority that you're planning to interact with because at some point in time, that's another factor. But yeah. I think that it does actually fit together all quite nicely.

DR. DWORKIN: I can't see that far back, though. It looks like it might be Lee Simon.

DR. SIMON: Yes. So I'd like to take issue with Nat and his initial comments regarding the regulatory quality issue and the scientific quality issue and suggest, as does the cartoon with Mutt and Jeff, that in fact you have to think about this in the context of the -- and everybody in this room has heard this -- the totality of the evidence.

The reality is that if you have a sloppy site that has two-days-late SOPs, that might be a harbinger, the canary in the mine, of something much more complicated that actually then is a

Page 103

marker, an insignia of a bad site.

Now, it truly may not have a direct impact on the quality of the clinical trial itself that took place at that particular site. On the other hand, if you'd begin to accrue more and more of these subtle problems, you begin to think about this in the context of, that site is a problem, or that trial is a problem, or something else is going on. I'm sorry I had you come all the way back to that.

So I think it's really important to recognize. And having been on both sides of this table, I can tell you it's really difficult to separate out what these issues are, particularly the ones that are the check-box issues. They're there because it ensures that somebody is thinking about those issues, while somebody else is thinking about the actual metric issues. And putting it all together, the totality of the evidence might make you be incredibly uncomfortable about what is going on with that particular data set.

So I think it's a little unfair to suggest

Page 104

that they were looking in the quarter two blocks away from what really may be important. In fact, all of it's important because it reflects the quality of the site or the quality of the system.

On the other hand, I think you've pointed out exactly what we need to do, which is actually real-time metrics, now that we can, of actually following this and really improving it.

I can't tell you the number of trials that I saw while in Washington and vice versa, what I now see as a clinical consultant. And it's pretty staggering how we actually get anything done and get approved, given the inadequacies of the monitoring systems.

DR. KATZ: Maybe I'll make one comment about that. I actually agree with that. I recognize, when I made that slide, that there's a danger that people are going to interpret it as being dismissive of the regulatory approaches. And the opposite. As I mentioned, I think those things are very important, and they can be harbingers of real issues at those sites.

Page 105

My point is more that are they direct or indirect? Are they close to what we're looking at or further away? And if you want to know whether pain is being measured accurately in a study, how does your SOPs being two days out of data, which certainly could be a harbinger of a problem, compare to the patient who doesn't know how to use the pain scale?

Which one is close and which one is far away? And my only suggestion is that we refocus on what's close to what it is that we're trying to produce and not be satisfied with things that are important, but are further away.

DR. DWORKIN: Let's go from the front back. John?

JOHN: This has been a wonderful set of talks, and I think the conversation so far is very enlightening. And what it reminds me of is that in thinking about the study of studies, when we teach about it, we talk about different kinds of error.

We talk about random error, which is, shit happens. There is confounding, which in an

Page 106

observational study are things that we hopefully can monitor, and understand, and do something about. And then there's bias, where the understanding is basically that there is a problem, that is not fixable with the analysis that we're going to do.

Now, what's been proposed here today, I think actually contributes substantially to that. And to, in a sense, complicate things before we try and make it simpler, we need to think about the kinds of error that we're looking at, number one.

So I would argue that as we go through the next couple of days, we may not be able to identify all of those, but we at least put some of the things we talk about into buckets to sort of say, is this something we can monitor and fix as we go along? Is this something that ought to be designed out from the beginning because there's nothing we can do about it if it actually happens? And then the last one being, what should we deal with in terms of the statistical analysis, the sort of random error that happens?

Page 107

The second issue that's implied by Nat's talk is that engineers also come up with a principle called limits. All right? There's a limit to which we can analyze and look at any problem. And when you're designing an airplane engine, you need to have microscopic limits with regards to how things change because those parts are going very fast, and if they move a little bit, it'll blow up.

If you're dealing with things that are much less complicated, your limits can be much larger. And one of the concepts that we ought to consider as we go through this next couple of days, too, is what are the limits here. And I would remind you that one of the primary limits we use, 0.05, was simply chosen out of the air one day.

So we ought to try and be a little bit more specific about some of these limits. And I'd be very interested to hear Nat's comment on how you decided when something turned red, because that's not obvious.

DR. KATZ: How do we decide when something

Page 108

turns red? It's not obvious. There are different types of signals that you can monitor for, and there are different approaches to determining what an appropriate threshold is for flagging them.

So we divide them into three categories. One is what I would call sentinel signals, or if one things happens, you want to know about it and it's not a statistical issue at all, so that might be an SAE or a major protocol violation.

Then there are things that we would call threshold limits where you decide arbitrarily that you think something going beyond a certain rate is likely to be problematic, so something like that might be in compliance with diary entries. If you're getting less than 80 percent of your diary entries filled out in a week, you decide that you want to know about that, whether or not it's a statistically significant deviation from your historical values, as is the paradigm for the control charts.

Then the third approach is a statistical approach where if a change in a variable, compared

Page 109

to its historical value, is very unlikely to be due to chance -- let's say it's three standard errors beyond the mean or whatever -- then it will flag an alarm.

Now, you don't always know how to interpret that. And it's very difficult to know what are false-positive and false-negative signaling rates are going to be when you used statistical thresholds like that. But that's kind of all you got in terms of a way of flagging things where you don't really have a rationale for drawing a particular line in the sand just based on what you know.

DR. DWORKIN: Ian?

DR. GILRON: Thanks, Sharon and Nat, for two super talks. I got a little concerned now when you sort of described the study patient as a measurement tool, which I understand that they are, but they also have another important role. They are representatives of all the patients who are going to receive this treatment if the evidence supports it.

Page 110

I'm sorry. I had a little flashback from our Bethesda meeting about assay sensitivity. And I just wondered whether we should make a distinction that when we talk about data quality and pain, we're dealing likely, almost exclusively with the primary outcome measure that's going to be based on self-report.

So if we were doing a treatment trial for sepsis, and mortality was the outcome, we might be tempted to sort of exclude patients who aren't going to die no matter what you do or are going to die very easily, no matter what you do, and exclude them from the sepsis treatment trial.

So getting back to the story of people who rate their pain as 10 or high variability, I suppose we have to be careful how we define a low-quality patient. So I mean, if they just can't understand the paradigm of pain intensity measurement, that's low quality. But if they're high variability, how do you designate them as low quality or they just have high variability?

DR. DWORKIN: So Nat has an editorial in, I

Page 111

think, this month's issue of Pain, that's relevant to your questioning. And so go ahead. Nat, you're prepared, I think, to address this.

DR. KATZ: Well, to make a long story short, the person is the measurement instrument. We can't get away from that. And it's just, we need to be honest about that, the concept of human beings as measurement instruments. The concept of having to rely on an individual's ability to look within themselves and estimate the intensity of these subjective experiences that they're having in research is an old idea, you know, back to the 1950s with Stevens, how bright is that light? How loud is that noise?

There are many different types of research where the human being is relied upon as being the measurement instrument. And like with every other human skill, different people have different capabilities at doing that well.

So I think I'll spare you all a long lecture on how one does that, but simply to agree with your comment that in classifying someone as not being a

Page 112

good instrument or being a good instrument, we have to be confident that that's not also biasing the results of the experiment in some way that would compromise our ability to generalize its results, that that is an important check box that you have to do before you deem someone to be good at reporting pain or bad at reporting pain. But how we do that, I think we'll leave that for the coffee break.

DR. DWORKIN: Neil?

DR. SINGLA: So thanks, Nat, for your talk. That was great, and both talks were excellent. I'm trying to reconcile and maybe put into context one issue regarding a human being versus a machine, which is that you described in your talk the point where it's above a quality control and the engineers go down under the floor and flip a switch or do something.

But the issue obviously in clinical trials is that it's not a machine that you can go down and flip the switch. You're getting data from real people. And as an investigator, oftentimes on the

Page 113

other end of that red dot, being the red dot -- meaning that you have a patient who started at an 8 and goes down to a 3 in an acute pain trial, and tells you they have no relief.

You're there. You're standing in front of the patient, and they're giving you data that would show up on a graph like that. You know it's discordant data. The question is what do you do? How do you educate sites?

I go around and educate sites about this. And they ask me, "Well, what am I supposed to do? I mean, the patient gives me this data. What am I supposed to do?"

I think there's been a lot of discussion of what should you do. Should you give the patient another chance to answer? Should you do nothing at all because that skews the quality of the data? Should you reeducate the subject on what the scale means? And we need to think about what is the right thing to do.

DR. HERTZ: Well, isn't that part of training patients ahead of time, having them

Page 114

understand things ahead of time? You don't want to be influencing a subject's response after the treatment has been delivered. You want them to report what they're going to report. But they need to have the appropriate skillset to use the instruments so that if there is something discordant, it potentially means there's something going on.

So maybe it's because their pain is low, but they're so constipated that they can't bear to move, so overall their relief is not great. I mean, there may be reasons for it. I mean, it's why we ask more than one question, because we're not always looking for concordant answers. We're also looking at the big picture.

So I think the answer is that, if the time is spent preparing subjects, training them so that they do understand what's being asked of them, then the results are the results.

So in that case, a red signal going off means you might want to look at it and think about it, but no. I don't think we should be intervening

Page 115

at that point. It's too late.

DR. DWORKIN: So I think this is a key question, Sharon. So let me imagine one of Neil's concerns. Let's say it turns out that patient AB hasn't been consistently filling out their pain diaries. Can the site -- patients now in the double-blind phase, and are not consistent about completing their pain diaries.. Can the site call up that patient on the phone and say, "Hey, you haven't been filling out these pain diaries."

DR. HERTZ: Well, that's different than questioning the nature of the response to the diary.

DR. DWORKIN: Okay. So there are some interventions that would be okay.

DR. HERTZ: Right, right, but that's very different to say you're not following the protocol. You're violating the protocol by not doing this.

DR. DWORKIN: So I'm going to pursue a little bit. What if the patient this past week --

DR. HERTZ: You're going to get to the point where I can't answer, you know?

Page 116

DR. DWORKIN: -- on a couple of days reported that their worst pain was less than their average pain? Is it permissible for someone from the site to call up the patient and say to the patient, "Hey, remember when we did the training before you started participation in this trial? We told you that worst means the worst pain you can imagine. An average is kind of your usual, but on Tuesday, Wednesday, and Thursday, you said that worst was less than average." Would that be permissible?

DR. HERTZ: I don't know because there, you're almost asking them to change the response. I think what could be permissible is to say, "I'm just going to refresh you on what these scales mean," and not make reference to data. I mean, it's a fine point, but if you say -- as part of your protocol, that if we start getting results as part of our QC-ing in real time that aren't necessarily consistent with an understanding or remembrance of instructions. separate from results, maybe they should be reported for everybody. So I

Page 117

mean, that's the sort of thing.
DR. DWORKIN: It wouldn't change anything, but we might do a little bit of a kind of brush-up.
DR. HERTZ: Yes. But I think it's extremely important to plan that. I mean, if you think you're going to have a population and the study is long enough where you may need to refresh them, I think just refreshing people who are giving responses that are potentially problematic is not giving everybody the same experience.
So if you think that's a risk, perhaps the approach that would be kosher is to say every two weeks, we're going to reinforce everybody and not just the people who are giving us results we don't think are helpful.
MALE SPEAKER: Say, Bob, can I just do a follow-up on this particular?
DR. DWORKIN: Yes. Jim Witter has been waiting. I'll come right back to you. Jim Witter has been waiting very patiently to say something about this.
DR. WITTER: I have a question. As we all

Page 118

know, there are many sources, causes of pain. We all also age, all of us. So as we age, we tend to get more comorbidities. So the question I have is the issue of attribution. When we ask patients to attribute the cause of their pain to a specific disease, if they have more than one disease, what are your thoughts on that?
DR. DWORKIN: The panel, not me.
MALE SPEAKER: They usually get excluded if they have another problem of equal or greater magnitude than the one that you're trying to study. So it has to be very clear that the subject can differentiate the pain problem that you're interested in from any other contributor to pain.
DR. DWORKIN: So that would be part of the training.
MALE SPEAKER: Well, it's part of the inclusion/exclusion criteria, as well as one of the things you go through in subject orientation and training.
DR. DWORKIN: Jim?
DR. WITTER: So this is kind of staying down

Page 119

in the weeds here on this specific issue that you raised, Bob. So with electronic data collection, a particular solution might be, in this case of worst pain/average pain, to the IVRS, may not allow an answer that is an illogical answer.
DR. HERTZ: That's not acceptable. Let me just interrupt. You cannot have your data collection impact the selection within a scale range. If your IVRS is set up so that worst pain cannot be less than average pain, you are changing the reporting of an individual, or disallowing it, or creating missing data because the numbers don't work. That should never be the case.
It wouldn't be the case on paper. It should never be the case just because it's electronic. That's hugely problematic.
DR. DWORKIN: Nat, Mike, would you like to add something?
DR. ROWBOTHAM: Well, I can add something that Nat actually showed a nice graph of. And that is, if you ask essentially the same question two different ways, using different scales, but it's

Page 120

essentially the same question, then you always have this issue of scattered, where a patient rates himself at low pain on one scale and high pain on the other scale, even though both scales asked you the same question.
So then you decide, do you eliminate answers outside of a certain boundary or do you use it as a composite measure?
I think it's an issue in design, having scales where, when the patient answers, you can get internally inconsistent answers, like average pain being higher than worst pain. That's probably an example of maybe not having the right set of measures in your case report forms.
DR. HERTZ: I mean, I recently filled out a completely unrelated non-health survey. It was about a shopping experience.
(Laughter.)
DR. HERTZ: And they gave me a list of questions with a yes or no. But something they asked me was not part of my experience, so it wasn't yes or no. It was not applicable. So do I

Page 121

say I was satisfied with it do I say it wasn't satisfied with it when I had no experience with it?
That made me irritated, and I cancelled the survey. Now they're not getting my opinion. So they created missing data by poorly wording their questions because I couldn't answer it. And if I'm a patient and you won't let me put in the number that I want to put in, that's frustrating, and now what do I do? Do I make it up?
It's not going to be relevant because I say my worst pain was a 3, but I'm only allowed to put 5 and up. So do I make it a 5 so I can get something done, because it won't let me go to the next page? What do I do? Or do I just forget about today? Then they're going to call me and tell me I didn't answer my -- I mean, it's creating a whole series of problems. So, yeah.
DR. DWORKIN: Does anyone else want to ask a question that can be definitively answered?
(Laughter.)
DR. DWORKIN: This makes chairing a meeting really easy. We were going up this row. We will

Page 122

get to the next row soon. Penney?
MS. COWAN: I just have a question. And I keep hearing the whole measure of pain, but there's so much more involved in pain itself, the function, quality of life, all of those things. And I mean, I'm not sure what all you measure in clinical -- I mean, we've talked about it. We just did a paper on physical function.
Are there other things that can compare that so that you can look at the pain score and make more sense of it?
DR. KATZ: Could you repeat the question, Bob?
DR. DWORKIN: Well, as I understood, Penney, you were saying that we've focused for the last 15 minutes on pain, but in this kind of identification, sources of error, threats to quality, is there any value in thinking about the other domains of the pain experience like physical function and mood? Can that contribute to increasing quality?
MS. COWAN: Right. And to clarify the whole

Page 123

question of the pain score, I mean, it would correlate.
DR. HERTZ: Absolutely. I mean, we're talking about how to sort of monitor inconsistencies or problems, but in terms of understanding the data, that's critical. And you're going to now call me a hypocrite because I don't allow what I'm about to say in labeling because it doesn't fit with our policy.
But when we look at data, we're trying to sort out what's the effect of this product in this population. And we look at the averages, and they're informative, but then we do spend a fair amount of time looking at different groupings, different individuals. We look at outliers. We do a lot of this to see what is the full picture.
If we have something that appears potentially inconsistent, we need to look. So is pain intensity shrinking and physical dysfunction worsening? Is pain improving but satisfaction is -- we do. It gives it context. And I think that was part of one of the earliest meetings on

Page 124

what are the relevant domains.
So I think that's all critically important and may speak to part of what's going on with some of this. This one particular example of basically a patient getting it wrong, it's really not a logical set of answers, is probably more one of understanding or the interface with the measurements. But, yeah. In terms of giving context, I think all these other parameters are very important.
MS. COWAN: I guess I wonder, is it a standard -- I mean, should they be standard measures in all clinical trials so that you can have maybe better outcomes, better measures of the pain if it becomes a standard. And I don't know if that's required in all clinical trials. It just makes sense to me that it might be.
DR. HERTZ: Well, rather than discuss requirements, which I will not discuss, the scientific concept underlying the six -- was it six or five? I always miss one. Five, then I would have missed two.

Page 125

The relevant clinical endpoints for study, that we're having a debate here, five or six --
(Laughter.)
DR. HERTZ: -- those core endpoints that were recommended as a group consensus for pain studies are still relevant. And for the most part, I'll say, just based on my experience, it's rare that we only get a measure. We get lots of secondary measures because they are useful for exactly that purpose, Penney.
MS. COWAN: I just wasn't --
DR. DWORKIN: So let's see if I got this right. Pain, physical function, emotional function, some patient global measure of improvement satisfaction, adverse events, and disposition in the trial; if they dropped out, why?
I've been favoring, from my perspective, the right side of the room. Let's see if there are questions on the left side of the room. We'll come back, Rob.
Anyone on the left side, all the way in the back, Dave Hewitt, and then Roy?

Page 126

DR. HEWITT: Yes. I wanted to ask Nat a little bit about -- because I am fascinated by the application of some of these other methodologies from other fields into our own. And I wonder if you have put much thought to something that's very popular in corporations these days, which is the Six Sigma process, which seems to me you're kind of skirting around a little bit. Mike was kind of talking about it as well.
Can you talk a little bit more about whether you've actually thought about applying Six Sigma processes to clinical trials and what would that mean? What would that look like?
DR. KATZ: Right. This is it. Six sigma is just a fancy buzzword for statistical process control methods in which things are flagged as being aberrant when they're six standard deviations beyond the mean. It might be six standard errors. I might have gotten that statistic wrong. And so that's become a buzzword, and there's a whole industry built around that. But it's just a particular way of thinking about the use of

Page 127

statistical process control to monitor processes.
Those processes could be your employment practices, what time people show up for work, what SOPs that you follow. It can be applied to anything. That's all that Six Sigma is. And the statistical process control principles developed by Shewhart are the foundation for that.
DR. DWORKIN: Roy?
DR. FREEMAN: Yes. So I'm struggling a little defining meaningful qualitative aberrations and how you do that. And under that heading, let me ask and pick on Bob Dworkin's example of logical inconsistency.
So you find a logical inconsistency in real time. Clearly, to reeducate that individual subject about the logical inconsistency is going to introduce some degree of bias in the study. So one potential scenario that I thought of is that you can prespecify that if there are X number of logical inconsistencies in a specific measure or in your scales, then you can reeducate the entire sample as to what the specifics of the measures

Page 128

are.
But even there, it's going to be hard to homogenize that because your clinical trial is dynamic. You could have 20 percent of your subjects in the trial at that point. But what about the other 80 percent who won't have that reeducation? And it's more actually, I think, a question for Sharon than it is a question for Nat. But Nat, as you were about to respond, why don't you go ahead?
DR. KATZ: Now, the analogy that I would give is you're doing a chemistry study and you've got five different pH meters in your study, and people are running samples through. And you find one of your pH meters -- you put distilled water in, which you're supposed to do once a week to make sure it's calibrated, and now it's reading a pH of 9 for distilled water, which we know it should be 7.
So what do you do? What you would do is you would recalibrate that instrument in order to prevent bias because it's your instrument that's

Page 129

off calibration that's introducing bias, not the corrective action to bring it back into calibration.

Now, back to human instruments, which is what we've been talking about, how do you know when it's off calibration and how do you know when you bring it back into calibration? How do you know when you're introducing bias? How do you know when you're correcting bias?

I think that's the point that you're bringing up. And if you're recalibrating that instrument, well, just the fact that you haven't done anything with your other instruments, does that introduce bias or are you actually confident that you're optimizing the issue of bias because you can read?

So these are the questions that you're talking about, and I think these are poseable questions, and these are answerable questions. And it would be shameful to address them on a policy level when they can be addressed at a scientific level.

Page 130

So I think, without getting into too much detail, you can actually measure the extent to which a patient is calibrated or not calibrated, and you can actually recalibrate them, and be confident that you've recalibrated them, and therefore prevented bias.

Now, I understand that there's a risk of doing that selectively and introducing problems, which is something that I think requires more discussion, but I think that we should focus on correcting bias and not prevent ourselves from doing that because we're worried that it's just a bad idea from a policy perspective.

MALE SPEAKER: Can I just speak to that for one second? Okay.

Quickly. Yes. I think the thing is it doesn't have to be that way. I think, in clinical trials, not only the patient but the investigator and their team should be reeducating. That should be a process of clinical trials over and over again, particularly for the controlled placebo effect as well as other measures.

Page 131

So I don't think it necessarily creates bias. You can introduce it across the line at certain intervals. I think it's very important because people sometimes, again, think of these patients as patients when they're actually study subjects and partners in this program, and they need to understand what they're doing. And so do the investigators, who may not understand that they are impacting the placebo effect.

MALE SPEAKER: I just have a question --

DR. DWORKIN: I'm sorry. Sharon, why don't you --

DR. HERTZ: I guess I would go back to a different question first, is how much of an issue is this? We're always going to get some scatter with regard to how people can retain instructions. It's normal.

If it is identified as a systematic threat to obtaining quality data that reflect the actual experience, then a systematic approach to dealing with it, like Dave just said, would be the way to do it.

Page 132

So I guess is it a problem? Is it a consistent problem? Is it enough of a problem to affect the quality of the study? And if so, the concept of quality by design is something that the chemists are using now in manufacturing. It's that sort of thing.

If you have an item here that you know is going to crop up in these studies, then design the study to address it at the beginning, so regular reeducation or updating of wherever the problem is to address. And then there's no bias because it's a protocol item. It's regardless of what's occurring. You're not reacting. You're planning and dealing with improving quality.

DR. DWORKIN: John Markman, then Mark Jensen, and then Mike McDermott, and then coffee, and apologies to all of you who had questions.

DR. MARKMAN: So I'd just like to come back to Nat's point about recalibration and tie it back to Ian's question, which as an investigator interacting with subjects every day, I find a real challenge.

Page 133

So we've seen a lot of research on baseline pain diaries, including or excluding patients who are perceived to be low quality, based on some of those results of those baseline pain diaries.

So the question I have is, should we be recalibrating at that level when we get a bad batch of baseline pain diaries? Should they be recalibrated then and then given a chance to reenter the trial once they've been reeducated?

Again, that goes to Ian's point because the real question here is, is pain a disorder -- which affects the instrument and the way you experience things? I mean, that's what allodynia is. That's what these problems generally are.

So again, it's hard for me to quite understand. I think this goes back to Ian's point. How much recalibration are you allowed to do of the rating instrument when that might be the underlying disease?

DR. HERTZ: But that's the question, isn't it? Sorry, Nat.

DR. KATZ: No. Go ahead.

Page 134

DR. HERTZ: Is it the underlying disease or is it the ability to use the instrument? So four patients who you really believe are having highly variable pain, they certainly should be studied because you want to know the therapy is going to work there.

But do you want to mix that in with a group that has less variable pain because, well, you have assay sensitivity to pick up an effect in anyone that way? So maybe the approach there is to take a population with one set of pain characteristics, and study it, and take a population with another if you believe that's the characteristic of the pain.

So I guess this is not telling you the important question of how you distinguish someone who has trouble with an instrument versus fluctuating actual experience, but if that can be sorted, then yes. They're all important.

DR. KATZ: Can I explain? Sorry, Mark. Maybe you were even about to say this. So variability consists of two components. Right? There's true variance, which is, my pain really is

Page 135

worse today than it was yesterday or worse now than it was this morning. And there's error variance, which is, my pain is really something, but then there's a certain error component put on top of that based on how good my primary endpoint measure is or how good I am at using that instrument.

So there's true variance and error variance. And the goal, I think we would all agree, is that we want to minimize error variance. And the question that everyone has been dancing around or talking about directly is how do we distinguish the two?

Briefly, I will just say that there are ways of distinguishing the two. For example -- and I'll repeat the same example we always give -- if you give the patient two different questionnaires that get at basically exactly the same problem, and they are widely discordant, there is evidence that that's an error measure. It's not like your pain can't really be high on one questionnaire and low on another questionnaire; it is what it is.

There are a number of other techniques that

Page 136

one can use to establish whether variance is error or true variance. And there are ways of training people to minimize and doing other things to minimize that error component.

DR. JENSEN: So there's been some discussion about training patients who have been identified to have a problem such as variability. Two other fixes. One would be, identify them at baseline, and it would be to exclude them from the trial a priori based on a protocol. And the other is to do these assessments after the trial is over and a priori say, if we identify patients who meet these characteristics, they are going to be excluded from the analyses.

So one is excluding patients and the other is excluding them from analyses. I'd be just curious, a quick vote from the panel, which of these would you recommend, none or others, assuming it's decided ahead of time?

DR. HERTZ: Well, what's the effect of that latter approach on the value of randomization and the intent-to-treat principle? So I see some

Page 137

potential problems with it from that perspective.
DR. KATZ: To that point, Mark, I don't have an answer to your question in terms of which is best, but we are doing a multicenter study now where patients are being tested for their ability to report pain accurately using experimental pain, where it has nothing to do with their clinical pain disorder. How well can you distinguish hot from less hot in terms of its painfulness? And those patients are getting excluded pre-randomization from the clinical trial.
So we're not excluding them based on the natural history of their own pain disorder. We're just excluding them based on whether they can actually perform that cognitive task well or not, again, pre-randomization so as not to violate that intention-to-treat principle. And whether that's better or not better than other ways, I don't know, but we're doing it.
DR. DWORKIN: Last question, a comment from Mike McDermott?
DR. MCDERMOTT: Yes. Nat, your definition

Page 138

of quality focused on minimizing sources of error that affect the accuracy of the treatment effect. I think you need to add the word "precision" in there. This is sort of a minor comment, but there's systematic and there's random error. And I think a lot of what these quality measures are going to accomplish is increasing the position, may be correcting some bias. And that's a different issue.
I was intrigued by the control chart discussion, and it's not just one process. You've got a process for every patient. But you've also got two different concepts of time here, and I didn't see the other concept of time come out so much as you got the concept of time where you've got patients being followed from the point of randomization.
You've got this other concept of calendar time. And I think the speakers later will be talking about detecting problems at the center level, not necessarily as patients are going forward in time, but as the center is going forward

Page 139

in time. And I didn't know if you had implemented any of those things in what you had your collaborator devise.
DR. KATZ: Yes, both. We have constructed control charts to look at site time. The sites started today, and they're moving forward, even though another site might have started six months ago. That was the nature of the control chart that I showed, which is site time. The date the first site begins, that's time zero, and it goes forward from there. And we also have control charts that look at calendar time.
So if for example things are changing between this year and last year, and how the study is performing according to those metrics, then that can raise signals as well. And I think we've all seen that happen in clinical trials, where you do an interim analysis or multiple interim analyses, and things look different based on calendar year. So we're looking at it both ways.
DR. DWORKIN: All right. We're going to take a coffee break, which is going to be outside,

Page 140

until 11:00 and resume then. I thank you all very much. I thank the speakers for a great start to the meeting.
(Applause.)
(Whereupon, a recess was taken.)
DR. PATEL: [In progress] -- medication adherence, and patient or participant misbehavior, and we have another set of great speakers lined up.
Mark Jensen is going to be speaking about pain reporting. Many of you know Mark. He is vice chair and professor of rehabilitation medicine at the University of Washington. As Dennis said, he is the editor of the Journal of Pain. He is a longtime contributor to ACCTION and IMMPACT activities, and he has published on a range of pain topics, including substantive contributions to pain intervention, behavioral intervention work, as well as methodological research on pain assessment and clinical trial methodologies.
So, Mark, why don't you go ahead?
Presentation – Mark Jensen
DR. JENSEN: Thanks, Kushang.

Page 141

So take a moment now and think back over the last 24 hours and consider what has been your average pain in the past 24 hours. Everyone come up with a number? Nobody came up with a number?

So it happens pretty quickly, doesn't it? It's pretty fast.

So the question is, was the number you came up with useful or valid? And I want to talk about that. I want to talk about the problem that can be associated with asking people to rate their pain, rate their average pain. What are the issues associated with that that can interfere with the validity of clinical trials? Talk about the two strategies that have already been discussed some to deal with that; patient training to improve their ability to validly rate their experience of pain, and monitoring how well patients are doing within the context of the clinical trial; talk about what we know about these strategies; and then talk about where we should go from here in terms of future research. And given what we know, what should we be doing now, which I think will be a topic of

Page 142

tomorrow afternoon's discussion.

So the problem is that people are not consistent. None of us are consisent with how we use these scales. We don't always comply with the procedures that are part of the clinical trial. And the bottom line is at this point, as a field, we don't know the impact of these inconsistencies.

How do we know that subjects are not consistent? Well, you've read the paper from Amanda Williams that 78 patients were asked to rate how bad their pain was on a VAS and a zero-to-10 scale, and there were problems, as she pointed out, with how people did that. The patients who had multiple pains -- of which most patients, in my experience, have multiple pains, it's the majority -- some rated their primary pain only at times. At other times, some rated only the pain that was worst at the time of the rating, not necessarily the primary pain of a study, and some combined them. Patients are inconsistent with how they use these scales and out of the context of training.

Page 143

A question could be are patients rating their intensity or how much they are bothered by the pain. Again, some patients reported that they think of the two separately. Some say that they are not able to distinguish between the two, and that it's just one big hurt.

The other thing that Amanda found was that there were 14 different factors that influenced the numbers that patients came up with, and these factors were inconsistent both between and within subjects. At some points, patients would rate pain differently as a function of different factors.

So, for example, when asked, "Do you consider how much pain impacts your functioning when you came up with your ratings," some said they often or always do. A large group said they sometimes do, and a group said rarely or never. Tiredness sometimes always or often, but also significant groups just did this sometimes. Sometimes people took into account their overall mood when they were rating their pain.

So the bottom line is that the numbers we

Page 144

get are influenced by a whole lot of factors other than just pain intensity, and that those factors differ across time for the same individuals.

This study was basically replicated by Joan Broderick and colleagues, in which they asked patients to rate the severity of their pain in the last week, and then they interviewed the patients about the strategies used.

Again, some people used the information from the entire week, but a large group didn't, even though they were asked to. Some generated an average, and a large group didn't. Sometimes patients considered their flare-ups, sometimes they didn't. Sometimes they considered times without pain, and sometimes they didn't. And sometimes the patients focused on just certain days rather than the entire week.

Most of the patients considered the impact of the pain, and that is a whole different domain other than pain intensity. We know that it is statistically distinct, and it is influenced by different factors. So if patients are considering

Page 145

the impact of the pain, you're measuring a different thing, even when you're asking people to rate their pain intensity.

So the issue, of course, is that we now know from these studies and others that pains [inaudible - mic fades] are inconsistent. What we don't know is the impact of this inconsistency.

Another way to say that is that you know from your own experience that to have pain is to have certainty, but as scientists, we know to hear about pain is to have doubt. We can say that we have considerable doubt about what we're actually measuring.

So how do we fix this? One strategy, as has been talked about, is to train patients before we even do the study how to use these measures. So, of course, the PROTECCT working group from the ACTTION, which is a part of the department acronym, which is a subgroup of UNCLE [ph] -- so Bob and Dennis are the men from UNCLE. I'm dating myself, I think.

(Laughter.)

Page 146

DR. JENSEN: -- is to train patients how to use these measures, a really good idea. How long has it taken us to come up with this idea? Decades. And in this particular strategy, you teach patients what is -- talk about what is their anchor for mild and worse pain. Teach them what average pain means, at least according to the training protocol; have them rate their primary pain condition only and not the multiple other pain problems that they might be experiencing.

Teach them to rate this distinct from their mood, fatigue, and the impact of the pain. Remember, 98 percent of people without training take the impact into account. And then say, "Listen, this study is about you. We're working together. We are research partners in this, and so it's very important, as a co-investigator of this study, that you help us get accurate outcomes."

Another strategy is developed by Nat, a brochure to teach patients how to rate their pain. And this one, again, elicit the cooperation as a research partner, introduce the zero-to-10 scale,

Page 147

have them identify from their personal experience what are the pain ratings associated with levels of 3, 6 and 9.

Tell them that zero pain and awake times only should be included for the average pain ratings, to take these into account, because we know that if you don't tell the patients what to take into account when calculating the average, some will take in the worst pain, average their worst pains. Others will just say an average of when I feel pain, and others will take it into account. So specifying that. And emphasizing the importance of accuracy, consistency, and that change can happen -- you don't have to always give a level of 8 -- and specificity. And then give them five examples to practice.

So we have these two training programs in place ready to go. They are currently being used even in some studies.

What are the unresolved issues? We still don't know if they improve things or not. What is surprising perhaps, given how much problems there

Page 148

are in understanding what patients mean when they give us a number, is that we are able to detect effects at all. The darn things work. The problem, of course, is that they may not work well enough. So we don't know if these training programs have any benefits, but research is underway. So we're starting to study that, which is a really good thing.

The other issue is that this inconsistency might be a trait. It may be, as Nat has alluded to, that there are patients who just aren't good at this, and all the training in the world won't help.

So perhaps the thing to do, if it's allowed, is to identify patients who are good at it as part of eligibility criteria as a way to make trials more efficient. We just don't know.

It's also the case that we have these training programs available. They may not get at what's important. We don't know that. These two programs that are currently available, they don't have 100 percent overlap, and there may be things that neither can get at. And, of course, training

Page 149

has a cost. It costs patient time, it costs investigator time to do this training, and, again, if they don't have a benefit, then that might be time wasted, or it might be simpler to simply exclude patients from the get-go who are bad at this. And there may be strategies other than training to get at increased reliability, for example, measuring a domain other than average pain something easier to recall.

So there's a lot that we don't know about these issues in terms of training.

In terms of monitoring, good idea. Nat introduced some ideas for how to develop systems for monitoring, and it seems like it's a useful way to monitor how things are going to fix problems.

What do we know about the efficacy of such approaches? Well, I did a scoping review about a month ago, and I did search terms. You search for pain and assessment monitoring and set the limits for clinical trials. The result of my scoping review was no studies. I'm not aware of any studies that have looked at this.

Page 150

So what do we now know about the benefits of monitoring on clinical trials I think it can be summarized in these ten or so words. Zilch, butkis, zip, diddlysquat, nix, nada, not, nothing, zero, not a bit, and nil. We don't know. It seems like a good idea, but we just simply don't know.

If we were to do it, what might it look like? These are things to consider if we develop monitoring programs. So certainly we want to ensure that the ratings, when we're asking for multiple ratings, are provided at the correct times.

How good are paper-and-pencil diaries for this? What do you think? No good. I think the field has pretty much said let's just simply not use paper-and-pencil diaries, to just stop it, and I think, in general, we have.

But there are a large variety of other options. You can page patients, use interactive voice, IVR assessments, have a person call the patient to interview, do it at clinical visits. In the clinical trials that I'm responsible for, I

Page 151

really like having a person call a person. That has worked out very well for us, and I've grown to trust it.

But we haven't done the head-to-head comparisons yet to see how these compare, which produces the best results. We simply don't know.

We want to make sure that the ratings that patients give us are consistent with the protocol and consistent with what you'd expect if they were providing valid and reliable measures.

So certainly, we could monitor the extent to which the average is between least and worst. But then the question is when we discover that's a problem, what do we do with it. I think useful discussions already have happened around this issue, and it seemed like there was already moving towards a consensus to say you ought not to just intervene when you see a problem.

So you don't want your interviewer -- when a patient says "My average pain is lower than my least," you don't want the interviewer to go, "Really? Really? Are you sure about that?" So

Page 152

maybe some kind of ongoing training. But it would be very useful to monitor this in all clinical trials so that we know as a field how often do we need to do the retraining, because we simply don't know.

We want to have at least some variability. If every patient always has a pain level of 7 before and after treatment, that's a problem probably. But as we've talked about there could be a problem with too much variability in terms of being able to detect a treatment response.

Then once we detect too much variability, what do we do about it? What are the other things that ought to be monitored? And I think that one of the useful things of this meeting might be to come up with a list of things that ought to be monitored in clinical trials, so that as we do more research, we can see which of those are most important.

So this other question, I want to plant the seed to start to think about what should be monitored on an ongoing basis in a clinical trial

Page 153

to ensure quality and to assess quality, because I think we simply don't know.

Everyone here probably knows about the studies on the impact of variability on ability to detect treatment effects. First study, Harris, with 125 subjects with fibromyalgia, and they compared placebo with milnacipran. They assessed current pain intensity 4 times a day for 15 days at baseline, and they wanted to know how much variability there was in these pain levels and was that stable. Are there people who just seem to have variable pain versus those who seem to have less variability? Is it a trait or is it a state? And what was the association between this variability and ability to detect treatment response?

They found that variability in assessment is quite strong. There are people who just report variable pain. Why is the question. How are they different than those who report more stable pain? What are the factors that contribute to more variability?

Page 154

Importantly, variability was moderately to strongly associated with response to placebo. Those individuals with more variable pain are more likely to respond to placebo.

The association between variability and active treatment response is very, very weak. Why? Very, very interesting question. And given this, of course, you'd expect that if you include patients who report highly variable pain in your trial, you'll be less likely to be able to detect a treatment effect.

John and colleagues looked at this in a very large study with -- they included over 2700 patients in 12 clinical trials, postherpetic neuralgia and diabetic painful neuropathy. And these were trials that specifically compared placebo with active treatment, 7 days of pain intensity ratings at baseline, and the question was, again, does pain variability predict the ability to detect a treatment effect as a sensitivity.

Indeed, as predicted, it was associated with

Page 155

ability to detect a treatment effect; that the more variability, a greater response to placebo of an active treatment, replicating Harris' findings.

Interestingly, the effects were stronger for postherpetic neuralgia than diabetic painful neuropathy. So there might be some effect of pain type that influences this variability.

My working hypothesis is that the variability may be related to pain problems that are influenced by multiple factors, not just nociception; things like hope, things like self-management, maybe perhaps centralized versus more peripheral pain. But we don't know the answer to that question. But it certainly seems to be associated with pain type.

There were meaningful significant effects, but perhaps not clinically meaningful for age for postherpetic neuralgia and diabetic neuropathy in terms of ability to detect treatment effects. But this variability in response in pain at baseline was a critical factor.

So given what we know, which is, I think,

Page 156

this, versus what we don't know, which I think is more like this, what do we do? We have very little to base our information on. I think clearly what we do is more research. I could wear a tee shirt that says more research is needed.

Some of the questions I think we need to address are: what are the patterns of inconsistency in responding to these measures that are most closely associated with assay sensitivity? We need to understand at least what they are, if not why. And which of these are modifiable? Can we actually train people to respond to these measures in ways that increase validity, or is it just certain people ought to be excluded from trials when the goal is to detect a treatment effect?

Of course, if you do exclude patients from trials based on something like variability, does that mean that you can't make conclusions about the efficacy of the treatment in that group? Does that limit your generalizability? So there are costs to doing that.

My leaning right now is that I think it's

Page 157

fair to a priori exclude some patients from trials who just don't know how to use these measures or have problems with them.

Which training programs are most effective or are more effective to changing modifiable factors? What is the effect of this training on variability? We can do this training, and maybe it has zero benefit. We simply don't know.

It's hard to imagine how it could make things worse, but I suppose that's possible, too. And are some pain domains easier to rate than others? And perhaps we might have a better ability to detect treatment effects if we measure other domains other than average pain intensity. Maybe training isn't needed. So these are questions.

Again, what I hope is that one of the outcomes of this meeting is that we'll come up with a list of the critical research questions the field needs to answer to help guide those of us who are doing research to say what are the key questions that we need to know in order to, 10 years from now, look back and say we now understand much more

Page 158

of the most important things.

But as trialists, what do we do in the meantime based on so little information? I guess we can only base it on common sense. Should we be using training in our trials? I suspect so. Of course, we can't say for sure what those might be; we can only guess. But it seems like it's a useful thing to be doing.

Should we monitoring? I think so. But then the question is what should we be monitoring.

So those are the discussions I think we need to have. And, as I said, if we should be training and monitoring, what should they look like? What's going to be most efficient? We don't want to use things that cost patients a great deal of time, are involved, or are very, very expensive. But we want to do them that are good enough that they are going to potentially improve our ability to detect effects.

Is it okay -- I hear loud and clear that it's probably not okay after you collect the data to remove patients, because that deviates from the

Page 159

ITT principle, from your analyses, based on identifying patients with problems who are unable to respond appropriately to measures during the time of the study. But it does seem like it might be fair to identify patients who are more able to use these measures more consistently ahead of time based on baseline measures and potentially use that as an inclusion/exclusion criteria.

But I think answering that question is an important issue for this meeting. I'd like to walk out of this meeting knowing an answer to that particular question. And that's all I have to say so far.

(Applause.)

DR. PATEL: Thanks, Mark. We have time for a couple of questions. Jim?

JIM: Thanks for a great talk, Mark.

The more we talk about this sort of people who don't get it, I guess one question -- I mean, we've all been thinking about pain for many years. And so I guess the question is, what is the problem with the majority of people who need training to

Page 160

fill out a zero-to-10 numerical rating scale?

So I can imagine a few things. They failed grade 8 math and haven't gone on since then, or all of a sudden, I thought -- and pain and cognition is really important, and left-brain/right-brain people. So what would the pain rating be like in a group of electrical engineers versus people who paint? And those are the people who would draw you a picture if you asked them to rate their pain intensity rather than that.

So I'm wondering whether the yield of training might be limited just because we kind of think most people should get it, and maybe the people that need the training may never get it.

DR. JENSEN: Yes. We simply don't know. I love to speculate, as any of you who know me know, so I'm going to do a speculation here.

If you've read Daniel Kahneman's book Fast Thinking-Slow Thinking, you know that there is a system for making quick judgments. If somebody walks in the room, you know you'll like them, you'll know if they're mad, you'll know if they're

Page 161

happy, immediately without thinking. Associate of learning, it's no thinking versus slow thinking when you process, think average.

I suspect that when you ask somebody to rate their pain, they use the fast thinking. They just use a quick judgment without really thinking it through. And when you go back and ask patients, "How did you come up with that decision," that's unfair because you're asking them to engage their slow thinking to determine how something so quickly.

If I said, "How do you decide within 2 seconds whether you liked Bob Dworkin?" Now, of course, you know you like Bob Dworkin within seconds, but how do you know that? How did you discern that? People go, "I don't know."

So it may not be fair to ask people how they came up with the judgment. I think it's a fast-thinking, immediate judgment. You just know. When I asked you at the beginning of the talk to rate your average pain, you came up with a number very quickly. And I suspect that many of you, many of

Page 162

us, don't know how we did that.

So it may not be that training helps because training is a slow-thinking thing. That said, perhaps teaching people to slow down, think about the past few days, do some "yeah," we might get a better number, we just don't know, and that's the issue, I think.

DR. PATEL: Okay. We have Crystal.

DR. CHEN: Yes. Hi. I have a question, being in pain for many years. But it's getting to the point, make me thinking, is it time to develop some other tools instead of a hundred percent relying on each individual subject's instant reacting to that question? Because we all know that it's quick, within 1 second, you know your number, you are coming up with. But that has many factors in there. Did you sleep well last night? Did you have coffee this morning? A lot of factors right there influencing that number.

This NRIS has become sort of a gold standard, however. I'm just challenged, and we are asking the question, is it time to develop some

Page 163

other type of tools? I can give an example in my previous life prior to pain.

I worked in restless leg syndrome, and there is a gold standard. It's called IRLS, and that is an interview. We know that RLS, restless leg syndrome, is pretty much measured by subject measures. However, they developed a scale, which is very powerful and very reliable, and the variability is under control; that is, the interview between the PI and the subject.

Of course, there is intensive training at the beginning to train a PI how to do the interview. We even have videos of a mock interview and not to lead the witness, per se, and not to insert any judgment. But if you do the training well, the study outcome is, I have to say, much more satisfying in many ways than just neuropathic pain studies.

That's just my challenge to our community and my question.

DR. JENSEN: So I think it's always useful to consider are we doing the best possible, and if

Page 164

somebody has a viable idea for a new way to measure it, that should be looked at and explored because it's always useful to have a better mousetrap.

That said, my current belief is that the training programs have been developed and maybe, perhaps tweaked and improved, might be a way to do that. This might be a way to train patients how to ask themselves, slow down and think about their experience, and rate what it is we're really after.

So that may be the new strategy, rather than say on a zero-to-10 scale, what's your average pain, say let's talk about what pain is for a little bit. Let's talk about what average pain is, and then do that for 20 minutes. And then given all that, as you think about -- slow down, think about your experience over the past three days, what number would you give it? I think that's very worth exploring in the next few years.

DR. PATEL: There will be more time to have additional questions. Thank you, Mark.

Our next speaker is Eric Devine. He's a professor and clinical psychologist at Boston

Page 165

University, treating addiction disorders, and is vice chair of the Boston University IRB, and he has carried out research on participant data fabrication. And one of his article was included in the background reading.

Presentation – Eric Devine

DR. DEVINE: So the data I'm going to present is kind of scary. As a clinical researcher, I look at this and say, what are we seeing in the literature if this is happening? And I don't mean to provoke fear among us, but maybe create some motivation to think about how can we design studies so that we don't have this problem in our studies, so we can eliminate some of the fraudulent data that happens on an individual subject level. That's the goal of talking about this.

Actually, maybe I'll give a couple of examples of why I ended up doing this research. I started out in clinical trials in the clinic I work as a therapist on a cocaine study. It was a NIDA program, a CREST rapid screening of all these

Page 166

agents that we had high hopes would help solve the cocaine epidemic. And as a therapist, I'd sit with patients, and they'd say, "Oh, so-and-so this in the trial, they're not here to actually change. They just want the reimbursement to buy more cocaine."

As a therapist, I'm sitting there, what can I do about that? It's hearsay, but it's a problem. And the answer to that question is find a way to not enroll people in the trial who aren't really there for the benefit of the trial.

Another example is an NIAAA-funded, multisite study. There were pretty good firewalls to keep individual identifiers out of the central database. I don't think it was truly de-identified with all 18 HIPAA identifiers out, but in this particular trial, there was an ancillary study that was added on where they collected identifiers. And lo and behold, one subject with the exact same identifiers at two different clinical centers, and their data did not match, not one bit, the primary outcome measure of drinking. There was no

Page 167

relationship between the two sets of data.

The last example I'll give before I show my data was an NIH-funded study to try and use motivational interviewing to reduce sexual risk behavior among people with serious mental illness.

In this study we had a screening visit, which paid $60, because it was a 4-hour visit, and we had a subject who had been through it who was selling access to our study for $20. She would say, "This is what you need to qualify, $20 kickback, and I'll tell you."

So she would send all these people, who actually never qualified. They were just there to go through the screening and they would rule out. But it was a one-time payday for them, and that's what they were after. And so in some ways, they were churning us for money the way a stockbroker would churn someone's portfolio just to make a quick buck in the day. Really troubling stuff.

So I went on to do this study of professional subjects, and professional subjects, as I define it, are people that enroll in clinical

Page 168

trials for the sole purpose of trying to generate income. And there is really clear evidence that these people are out there.

You've got subjects that participate in multiple trials. They report enrolling in different trials at the same time and lying about it. They use deceptive practices where they conceal information. They enroll in a trial. They never intend to take the medication, they don't take the medication. They pop it out into the trashcan.

We actually had an example of that brought to our attention in our NIAAA trial. The security camera in the parking lot observed someone over a trashcan with a blister card punching out, and they had just left their appointment. And that was great because we were able to exclude them from the study. But we don't often get that kind of eye in the sky to detect fraud. So we need other measures to do it.

So there is some research on professional subjects, but a lot of it is around the ethics of

Page 169

disproportionate risk and these people not having generalizable data. Not many people have actually gone out to say how much is this happening, how many of these subjects are actually doing this, and that was the goal of my study.

I recruited 100 subjects through news print and online advertisements, and I just billed it as an experienced subject study. So I just wanted people who had at least two studies in the past year and three in their lifetime. And I put the advertisement in the Boston Globe, the Boston Herald, and Craigslist, and the free paper, the Boston Metro.

These are all sources in the Boston area where we have a lot of funding. Everybody is using it. The Metro has a two-page ad for clinical studies every day and it's one-stop-shopping for someone who wants to make money.

This is an example of the ad just so you can see I wasn't looking for liars, cheats and scoundrels. I was just looking for someone who has been in multiple studies.

Page 170

So what I have, I have a questionnaire; actually, it's an interview. And I asked people about their rate of concealing information. So the kind of things I'm worried about in my clinical trials in alcohol and cocaine dependence, are they hiding health problems which could be a risk for them being in the trial based on the medication or mental health problems that could -- they might be hospitalized midway through the trial.

There are a lot of potential risks to the subject from concealing, but more important, there is a lot of risk to the integrity of the data because they may not tolerate the medication. We have an adverse effect profile that is really not consistent with the population because we don't understand the population and we don't understand the population characteristics.

So I actually asked them about this stuff in a pretty transparent way, and I have to say this data, if anything, might underestimate the rate of fabrication and concealment because these are people that like to fabricate and conceal, and

Page 171

they're in the study.

(Laughter.)

DR. DEVINE: So the fact that I even got this, I think this is the tip of the iceberg.

I also asked them about fabrication, which I think is a more egregious type of deception in research. Genuine subjects might conceal things because they are desperate for help and they want to get in, and I understand that, I accept that. It's not good. But fabrication is a different brand of deception, which is more worrisome, where people might pretend to have a health problem in order to enroll in a study or might lie about the very symptoms of the disease that I'm trying to study once they're in.

So this is above and beyond someone who is genuinely trying to get help. So we asked a series of questions. And just a little bit about the demographics in the sample, skewed a little towards the male side. Notice the income here. Most of the sample was below $30,000, which in Boston has got to be the poverty level. I know it's not

Page 172

nationally, but it's so expensive to live there, and an average of 12 studies in a year. And the majority of them have had experience enrolling in medication trials. All of this was in the paper that was circulated.

Let's look at what I actually found. Lifetime concealing. Did you ever conceal something in order to get into a trial? Seventy-five percent of these people said, "Yes, I did conceal." I don't know how often they do it, and I don't know what type of studies they do it. That's a whole other level of study, but I do know that they report doing it.

What are the most common types of things people conceal? Participation in another study, 43 percent; health problems; other prescribed medications. This is a big one. We really worry about this when we think about the adverse effect profile in a clinical trial, that it might be misattributed when it's some synergistic effect of medications or effect of a different medication.

Recreational drug use certainly is something

Page 173

we worry about in our alcohol trials, mental health, alcohol use, and legal issues. People sometimes disappear in our cocaine trials midway because a case is pending, and they don't want us to know that, and that's something that affects the overall integrity of the study.

On the fabrication scale, a little lower number in terms of the risk to study integrity; 33 percent of the subjects reported that they do fabricate information in order to gain entry into a trial. And among those different types, exaggerating symptoms of a disease, which I can see as something that genuine subjects might do. They really want to make sure that they qualify for an alcohol study, so they exaggerate their level of drinking. So I get that. It's not good for the data, because once they're in, then suddenly they're drinking a lot less. They're telling the truth, and it looks like there is a medication effect that really was just a return to maybe telling the truth.

Pretended to have a health problem,

Page 174

14 percent. Given research -- the symptoms are the focus of the study, false information, 12 percent. And this is a good one, done intentional harm to yourself in order to qualify for a study. People do this. People will hurt themselves so they qualify.

Two examples I can remember from the interviews. One is someone gaining a substantial amount of weight to have the right BMI for entry into a trial, and the other is someone who discontinued their antihypertensive medication so that their blood pressure would be out of control for entry into a trial. So stuff that we really don't want to see.

Then on the last one, tried to enroll in the same study twice using different names or changing identity. People do this.

There were other forms of deception, and I highlighted in yellow findings that I didn't publish in the article that was circulated before the meeting, and I think one of them is particularly egregious and worrisome. It's halfway

Page 175

down. Enrolling in a medication study where they have no intention of ever taking the medication, 11 percent of subjects reported doing that.

When you think about the dosing problems and optimal dose, this is the worst kind of medication compliance problem you can think of. Come up with a reason to stop taking the medication without losing reimbursement, adverse effects is what they use. And there again, you've got problems with the study design and also the safety profile.

There are a bunch of other things that people do. They share information, they give and receive it, and they tell us that they're getting better when they're not getting better.

There are also some open questions I asked about how do you game the system, basically. And one of the ways, they do study clinicaltrials.gov. They go to websites, and there are several organized professional subject websites where they can get information.

They will telephone screen in a group and share information. And my favorite of the

Page 176

strategies is the organized research kingpin, and that's actually the word they use to describe it. In the Boston area, there is someone at the VA doing this, at the JPVA, and there is someone at BMC. They are different people. And they sell access to a study.

So they have all the entrance criteria for screening in that you need to say. And for a fee, they'll give you that, and they'll tell you who to call. And if you don't give the kickback, you don't get any more information. It's an organized group of people that do this. And for me, as a researcher, it is frightening to have that.

So the goal is to figure out how do we prevent this, and I have worked really hard to keep professional subjects out of my alcohol trials, in particular. It's harder with the cocaine studies. And it has a lot to do with the study design and how I advertise.

For studies with direct benefit, I never include reimbursement in the advertisement. That's like just putting out a sign that says come and

Page 177

game me because I've got this money you can make quickly.

When I recruit people, I don't offer much money for the initial screening visit, if any at all, because some people will just come in, they know they're going to exclude from the study, but they just want to churn for a one-time payout. They're interested in that prorated payout, where they spend a few hours and they get $20, and that's enough to get by for the day.

Telephone screening, we spend a lot of time trying to make our screening process as non-leading as possible, not giving away the criterion along the way. And even once they come into baseline, if we can include more objective measures, we do that. A lot of open questions, a lot of converging information from different interviewers to try and see is this person really what they appear to be.

We'll look at the medical record. That's part of our consent, if they're getting care at Boston Medical Center. We've had people that deny any medication. And we open up their chart, and

Page 178

they're schizophrenic, and they've been taking antipsychotics for 20 years. But we'll do that because we really can't afford to have people enrolling in the trial who have that risk.

Certainly, we can spend time looking at subject motivation, and professional subjects are very focused on questions about reimbursement, when it's going to come, how much it's going to be, what will happen if they only do part of something, and maybe less focused on direct benefit of being in the trial or risk of the medication.

We do look at some inconsistent data. Someone comes in and is drinking 50 standard drinks a day and blowing a zero on the BAL, it's like that would really be hard to do. You just don't metabolize alcohol that fast.

So we're always alert to these sort of things, and I can talk more about this during the discussion panel later. But our goal is to really, as much as possible, use strategies to keep them out of the trial all together rather than trying to deal with it on the other side, where I think the

Page 179

statisticians are going to look at us and say, "Who is this professional subject that's introducing all this? I can't tell." It would be nice if they could. And that's it.

(Applause.)

DR. PATEL: Thank you, Eric. You don't have time for questions, unfortunately. So we'll have Bob Dworkin come up. Thanks a lot.

Bob is the herder-in-chief of the meeting, so I don't think he needs an introduction. But he's the founding and executive director of ACTTION, and he's going to be talking about patients.

Presentation – Robert Dworkin

DR. DWORKIN: Thanks, Kushang.

So my role is clear. You've all gotten indigestion from Eric's talk, and so for the next 15 minutes, I'm going to try and kind of reduce your indigestion before lunch because, of course, those data are really quite alarming.

So I want to just talk about a couple of strategies that might be effective in combating

Page 180

some of the misbehavior or misconduct that Eric has really described. And the primary strategy, and many of you have heard this before, is to blind everybody to everything they don't need to know, the site staff and thereby the [inaudible – mic fades], and to use blinding to a much greater extent than we typically think about it, which is double-blinding the treatment assignment.

So if you take nothing away from the next 15 minutes other than blind everybody to everything they don't need to know, that would be the most important takeaway message.

I think many people in the room are familiar with these data and data like these. This is a dramatic illustration of what in psychiatry is called baseline score inflation. Let's see, where is my pointer? Just in case you've ever seen this, it's almost as alarming or maybe more so than Eric's data.

So these are clinician-rated Hamilton depression scores in a clinical trial of some SSRI, probably for major depression, where the inclusion

Page 181

criteria -- I can't see the numbers -- were something like around 20. You only got into the clinical trial for major depression if you were baseline. Before randomization, the Hamilton score -- the patient was 20 or above. And so these are clinician-rated measures.

These are the patients' responses for how depressed they were. You can see there's no correlation between the clinician's pre-randomization assessment of depression and what the patient said their depression was at the screening visit.

So the kind of deduction from these data is that the clinician investigators at the sites were inflating the patients' ratings of depression in order to randomize them to enroll them in the clinical trial.

These are Hamilton scores at the end of the trial, where you can see there's a really nice correlation between the clinician ratings and the patient ratings, because, of course, at the end of the trial, the clinician doing the Hamilton ratings

Page 182

has no motivation to inflate the scores.

So data like these are disturbingly common in psychiatry trials and led to David Hewitt, who is in the back of the room, to propose this back in 2011, propose and implement this in a clinical trial.

David said in his clinical trial, both investigators and patient were blind to the following information: entry criteria for patients pain intensity, baseline pain intensity, definition of responder groups, visited, which randomization occurred, blah, blah, blah, blah, blah, blah.

So this is the first example in the pain literature that I know of, of blinding everybody to everything that they don't need to know. And if you have any questions about how we came up with this, David is here at the meeting.

We implemented this in a clinical trial -- Andrew Rice here -- of an angiotensin 2 receptor blocker that was published a few months ago. And I just wanted to show this slide to kind of make the point that there really aren't any

Page 183

problematic aspects to blinding people to all of these aspects of the study design.

There are a couple of things we can talk about later in terms of IRBs, in terms of cranky investigators who wanted to know all the trial details, but the bottom line is it's easy to blind sites to information that they have no need to know. Obviously, we're not blinding things that have implications for safety, et cetera.

So one potential recommendation is, the obvious, blind patients and site personnel to absolutely everything they don't need to know. Also, of course, as you saw from Eric's presentation, you want to do the same thing for clinicaltrials.gov, because there is no reason to give a site a redacted protocol if all the information is available on the Web. And so clinicaltrials.gov has to be blinded, if for no other reason, to prevent the study kingpins in Boston from educating pseudo patients how to participate in Eric's addiction clinical trials. So blind everybody is one possible recommendation.

Page 184

Now, this is Jeri Burtchell. She lives in East Palatka, Florida, and she has relapsing remitting MS. And a number of years ago, she participated in a double-dummy trial of IM interferon beta versus oral fingolimod for relapsing remitting MS. And she went home after her first randomization visit and started a blog about her experiences in the trial, where she explained to patients, other patients in the trial, on this blog, how it was she became convinced very quickly that she was randomized to oral fingolimod versus IM saline.

By the way, she was correct. She was, in fact, in the oral fingolimod group, and it had to do with the fact that when the nurse injected her thigh, she didn't feel anything. And in the past she had been treated with Avonex and the injections always felt a little bit of burning. Plus, she saw that her blood pressure dropped at one visit, which is an event associated with fingolimod.

So she had a blog where she was basically training patients around the country who were

Page 185

participating in this trial how to unblind themselves, which is a big problem.

Because of this blog, she was invited to a bunch of professional meetings over the next several years, and she became convinced that what she had been doing was terrible. And she now has this site. She took her blog down because she understood over time that what she was doing was really threatening the integrity, the quality of the data.

She now has this site called Partners in Research, where she is doing everything she can -- and it's a great site -- to kind of educate the patients that they need to be collaborators, partners in clinical trials, to encourage them to participate and to encourage them to do a good job when they are participating in the trial.

I only found out about all of this recently, Otherwise, we would have invited her to this meeting, because it would have been great to have her here.

So there is patient misbehavior, as Eric

Page 186

pointed out, and it is known and it has been known for a while. This is an Institute of Medicine report from 2010, where -- I hope you can read the slide -- the IOM report describes the existence of psychiatry trials of professional patients, as you've heard, noting the example of a 300-patient schizophrenia trial where 30 patients were participating in the same schizophrenia trial in multiple sites.

This is a schizophrenia trial where 10 percent of the patients in the trial were participating in the same trial at multiple sites, presumably for the compensation. And the response to this recognition of duplicative participation is there are now -- I don't know what to call them -- websites, organizations, outfits that provide registries where sponsors and sites upload demographic information and the -- this is one of them, DupCheck, which has been started by Jonathan Rabinowitz as part of the IMI effort in Europe.

What Rabinowitz does with DupCheck is to identify duplicate patients not only within trials,

Page 187

but between different trials, and then notifies the sponsor that a duplicate patient has been identified. These slides were sent to me by Jonathan. So he obviously has a conflict of interest.

This is interesting. I don't want to spend too much time on it. When patients are duplicate enrollers in the same trial, what is it they think? He's done a little bit of research on it. They believe they know better than the investigator, and think it's a silly requirement that you can't participate in the same trial at multiple sites.

They miscalculate how long ago it was that they participated in the same trial. They say they're not a criminal, there is nothing wrong with what they're doing. They want to get paid for an additional study.

So this is happening, and as Jonathan says, DupCheck can be used in one of two different ways at the time of screening: to exclude these individuals who are trying to game the system, or after the data have been collected -- prevention is

Page 188

better than treatment, but if you haven't excluded them in advance, to identify them and do something about it during the data analysis stage.

This is an example of a schizophrenia trial. I don't really know the details. He presented this at a schizophrenia meeting, as you can see from the bottom of the slide, where he identified a not very large number of duplicate patients in schizophrenia trial and shows that the significance of the trial before removing the duplicates was not significant, was .054. And then after he removed the 10 duplicate patients, the trial became statistically significant.

Take this for what you will, but, obviously, removing duplicate patients from a trial is a reasonable thing to try to do.

Mitchell Efros -- actually, no, this is Thomas Shiovitz, who runs a clinical site in Southern California, the Los Angeles area, and he has set up a network of CNS sites in Southern California that, in a HIPAA-compliant way, share patient identifiers.

Page 189

I hope you can read the title of the slide. He is using the subject registry to create a duplicate-free corridor for conducting clinical trials. So he's hoping that by linking together these CNS sites in Southern California, they've created a duplicate-free corridor because they're sharing information, allowing the duplicate patients to be identified.

He has also done some research on this question. And in this abstract, how far are duplicate patients willing to go, he's showing that most of them are willing to drive at least 25 to 50 miles to participate in multiple studies simultaneously. That was one meaning of how far are they willing to go.

Another meaning is how far are they willing to go in terms of varying their diagnosis, and the answer to that is these individuals, no problem being a schizophrenic on Monday in Los Angeles; and then Tuesday driving down to San Diego and having bipolar disorder for a bipolar disorder trial; and then on Wednesday in Irvine having generalized

Page 190

anxiety disorder.

So they're willing to go physically far, and they're willing to go diagnostically far in order to participate in multiple trials at the same time.

Mitchell Efros has a site that's called Verified Clinical Trials. This seems to be more in the U.S. as opposed to DupCheck, being more European. I don't really know anything about the relative value of either of these approaches or Shiovitz's Southern California network, but I think this is on the horizon because as Shiovitz says, and this is from an e-mail he sent me, while there is not 100 percent adoption of a system to prevent duplicate patients, the main message he wanted me to convey to you all is that use of some system is better than not using one at all. And it seems hard to disagree with that, to me.

If there are ways of potentially duplicating these -- potentially identifying these fraudulent, fabricating, duplicate patients, why not consider using these approaches?

So the second potential recommendation,

Page 191

implement efforts to identify and eliminate fraudulent and duplicate patients. And this goes back to Eric's presentation. Let's do our best to verify the patient has the disorder, has the symptoms, either by getting medical records from their primary care clinician or going to electronic health records and then consider using these networks, DupCheck, Verified Clinical Trials, Shiovitz's network, to identify duplications before they are randomized.

So the last thing I wanted to say, and to say I mentioned before, were we blinded everybody to everything they didn't need to know, we implemented -- and this is mentioned in the article. Andrew and I and the other investigators implemented a baseline pain exclusion algorithm. We kind of created an algorithm where we interrogated the baseline pain diaries, and based on the pattern of responses, excluded patients. And the implementation of that was kind of straightforward.

So what is an example of it? This is not

Page 192

the example we used, but it's an example of the kinds of things we have been talking about this morning in Mark's presentation, in Eric's presentation. And so one can imagine a completely blinded algorithm that is applied to a patient's week of diaries during the baseline prerandomization phase of the trial that is the basis for excluding some patients from participation.

So they have to complete at least six out of seven diaries. The mean of their ratings has to be between 5 and 9, not too mild, not too extreme. Some variability exclusion, a la what Harris and Farrar and other people have published.

This is NATC's point, looking during the baseline week at agreement with two different ways of assessing pain. It could be agreement between average and worse pain or NRS and VAS or, as Nat mentioned, a pain measure that's more generic and something disease-specific, like the WOMAC, one or more days with worse pain, less than average pain. They didn't pay attention to the training. They

Page 193

are not thinking through their ratings. So maybe if they mess up in this way, they should be excluded.

I went back and forth, and this is all just for argument's sake. One might give them a day of being a little sloppy, and you might want to make this greater than one sloppy day, but maybe not.

Then finally -- we're going to hear much more about this this afternoon -- some trials have a placebo run-in and maybe it would be reasonable to exclude patients who demonstrate poor adherence to taking the placebo during the placebo run-in, because if they're not adhering to the placebo in the placebo run-in, then isn't that a bad sign about what their adherence is going to be during the rest of the trial?

Of course, if one were to implement something like this as an exclusion algorithm prior to randomization, so this would be prior to randomization, one would hope it increases assay sensitivity. One would hope it excludes some fraudulent patients.

Page 194

My last slide, it has to be totally blinded. So I wanted to begin with the importance of blinding and end with the importance of blinding, because we want to do everything we can to defeat the study kingpins that Eric told us about.

So thank you very much. We won't take any questions now. This is time for the lunch break. And because we went a little bit over, let's reconvene here at 1:10, and there will be ample time for discussion after lunch.

Thank you all very much.

(Applause.)

(Whereupon, a luncheon recess was taken.)

Page 195

A F T E R N O O N S E S S I O N

MODERATOR: Let me introduce the next speaker. Dr. Bernard Vrigens is a biostatistician by training, and he's currently the chief scientific officer of MWV Healthcare, which is a packaging company. And he has a lot of experience with monitoring and evaluating medication adherence.

I had the pleasure of having dinner with him last night, and he told me that he's been back to the United States five times in the past nine weeks to give talks similar to the one that we're about to hear. So he's highly sought after, and we're looking forward to his talk.

Presentation – Bernard Vrijens

DR. VRIJENS: Good afternoon. Thank you very much. Thank you for the invitation to give me the opportunity to talk about medication adherence here today.

Today we have to deal with a lot of very effective therapies. But if we don't have appropriate adherence to medications, we will not

Page 196

achieve effective disease management. And in most of the adherence talks, you will see they always start with the sentence, "Drugs don't work in patients who do not take them." However, the sentence assumes that adherence is a dichotomic process, either you are adherent, either you are not adherent. And what I will do in the next few slides is to convince you that it's much more complex.

First of all, about the taxonomy. A few years ago, we were called by the EU, Union, to come with recommendation on adherence for the European Union, and we were sitting together with seven universities around the table, and we didn't know what we were talking about.

Was it adherence? Compliance? Persistence? Concordance? And all the translation in all the European languages, we didn't know what it was, really. So we defined it as a process by which patients take their medications prescribed, but we recognize that it is a dynamic process over time and that there are three key elements.

Page 197

Once there is a prescription, first the patient has to initiate the therapy. That's the first thing, initiating therapy. Once the prescription is initiated, the patient has to implement the dosing regimen, meaning taking it once a day, twice a day, with food, without food, and when we think about polypharmacy, this implementation piece can be very complex. And then the patient has to persist with treatment for a long period of time, especially in chronic diseases.

So what can go wrong in that process? Either the patient doesn't initiate, and that's a dichotomic outcome; it's yes/no. Either the patient delays, takes an extra dose, omit a dose, and that's a dosing history; it's a time series. Either the patient discontinue treatment too early, and that's a time to event.

Statistically speaking, those three elements are very different in nature, very different in nature. That's why we need to identify and to tackle them separately. And that's why, for

Page 198

example, if I say an adherence of 70 percent is very confusing because you don't know if the patient is really taking, implementing, 70 percent of -- taking 70 percent of the medication or stopping 30 percent too early. You don't know.

Okay. What are the measure that were have to measure patient adherence? And I have classified the measures here in four categories. First of all, on the lower end of this figure, you have methods that are biased.

For example, what we do mostly in clinical trials, pill counting, is extremely biased. I will show you some data afterwards. But people tend to drop the pills before showing up at a visit. And a retrospective questionnaire, asking the patient if he took the medication in the last month, typically is extremely biased as well.

On the upper part we have, for example, therapeutic drug monitoring, which is an extremely reliable method, but it's very sparse. Why? Because you have the idea of the adherence at the time of sampling, nothing before.

Page 199

Most of my research has been focused really on electronic monitoring, and the idea is that we put the chip in the package, so that every time the patients open the package, it's time-stamped, and we know when the medication is taken.

First slide. This method is really rich and reliable, and I will show you some data about that. My first slide here is about what is pre-electronic? Pre-electronic methods are unreliable. And one example is here, the first upper figure. I like to show this one because it's done by a very famous statistician in London, Stuart Pocock.

Rather than giving a hundred tablets to the patient for a hundred days, he give 160 tablets for a hundred days. So the patients were expected to bring back what is indicated here by the arrow, the blue arrow, and you see a very nice distribution around the blue arrow, but you see 20 percent of the patients bringing back an empty bottle.

Those typically are patients dropping, and this 20 percent comes always back; when you compare pill count with electronic monitoring, you come

Page 200

back to those 20 percent, and those are not at random. Those are the worst patients who drop the pills before showing up at the visit.

This is an example also about "white coat compliance" affecting therapeutic drug monitoring. This is an extreme case that we encounter. Patient takes a drug, takes one tablet, comes for the blood sampling. Takes nothing. Takes two tablets, comes for the blood sampling. Takes nothing. Takes three tablets, comes for the blood sampling. Very good. He looks okay. Well, he never took the medication, really. And we found bias in 31 percent of the samples clustered in 66 percent of the subjects.

Also, when you do self-report, really, it's sky-high reporting adherence compared to electronic monitoring. And multiple studies have shown that physicians or healthcare providers are very bad in predicting medication adherence.

So that's about bias. But the most important for me is really to see the dynamic in adherence, and I will show you some examples.

Page 201

This is an example of electronic monitoring. You can see this patient is on a twice-daily-dosing medication. On the X-axis, you have the follow-up in the study, on the Y-axis, you have a time of drug intake. Every blue dot, it's a dose taken.

So you can see this patient takes the morning dose at 7:00 every day, exactly at the same time, and he takes the evening dose exactly at 7:00 p.m. He was two minutes late here, and then he is perfect. And there is exactly 12 hours between the morning and the evening dose. So this patient exists.

But as you can imagine, a lot of patients do deviate from that perfect pattern. Something that we encounter very often is this type, where you see weekdays patients take the medication at 7:00 in the morning, but weekends he sleeps out and takes the medications about noon, but still a very good adherer; he never missed a single dose. Those are patients from phase 2 clinical trials, so they're real.

This patient, you can see every time there

Page 202

is a gray bar, it's a missed dose. So the morning dose is okay at the beginning, the evening dose is already problematic, and the further in the study the worse it is, especially at the end. In the evening, he almost takes nothing any more.

I like to show this patient also. This patient, as you can see, has problems at the beginning, so at initiation missed a lot of doses. Then he does it much better. And then there is a full stop of treatment. There is a full discontinuation of treatment about halfway. But this patient shows up at the last visit and doesn't tell he stopped medication before. He just goes for the last assessment, and he didn't take the medication for a while before.

I could show you thousands of those patients, but this summarizes the data from about 17,000 patients coming from 95 clinical studies. So what you see here is the blue curve. The blue curve gives you the proportion of patients who are persistent with the treatment, who are still engaged with the treatment. And we see that after

Page 203

a year, we have lost about 40 percent of the patients in clinical trials.

It's also striking to see that the day 1, we have a drop of 3 percent. So that means that even in clinical trials where patients are highly selected, patients are highly motivated, patients get the medication for free at the investigational site, 3 percent of them roll back home and never open the box.

Then we have here the adherence curve, the red one, that over time, it gives you the proportion of patients who open the box as prescribed every consecutive day. So if I take the hundred, I have 80 percent persistence but only 65 percent of them who do it as prescribed.

So that means that every day, because you see those two lines are pretty parallel -- that means that every consecutive day among the patients who are still engaged with the therapy, about 15 percent of them do not do it as prescribed.

Summary. After a year, we have lost 40 percent of the patients. Every day, 15 percent

Page 204

do not implement as prescribed.

Persistence is very different across diseases, and we see the worst persistence in CNS, mainly in depression studies. That's where we see the worst persistence. And you have to think about all the diseases that are associated to depression, like oncology, for example.

Two examples. One is hypertension. This is a subset of the database. I thought it's interesting because the persistence in hypertension is about 50 percent after one year, but implementation is slightly better. You only have 8 percent non-implementation. Why? Because this is a very easy to take medication. It's once a day. Extremely simple, so you have slightly better implementation, but after a year you have lost about half of the patients.

What is interesting from this study is that we looked at when they take their medication. This is missed doses against day of the week. So the patients who take it in the morning are the ones who miss the less doses.

Page 205

Then the ones who take it in the evening, they miss slightly more doses, especially on Saturday evening. And then the wobblers are the ones we cannot classify, the ones who take it today in the morning, tomorrow in the evening. They have no time patterns. Those are clearly the ones who miss most of the doses.

This figure is interesting also because it shows the persistence curves stratified by implementation. You see that the better you do it on a day-by-day basis, the longest you persist with treatment.

So that we come back later in my presentation when we are thinking about interventions, if we can work on building a habit in the patients to do it better on a day-by-day basis, we increase the likelihood that the patient will persist longer.

Everybody will tell me, in our research, we started in hypertension, and then we did diabetes, and then we did -- and every field always, oh, yes, that's hypertension. We know that adherence is bad

Page 206

in hypertension. And then you go to diabetes. Oh, yes, the diabetes is the same. And then you go HIV. Oh, yes, HIV is also special. So field after field, we see major issues in adherence across all therapeutic areas.

I want to take this opportunity to show you this publication that was out last week -- no, last month, sorry -- from Alabama University. They've randomized -- no, they didn't randomize -- they have followed 500 kids with leukemia. They have followed 500 kids using those electronic monitors.

They realized that the kids who had an adherence, global adherence, above 95 percent, which requires very precise implementation of the dosing regimen, had a 5 percent relapse rate. And the kids with lower than 95 percent adherence had almost a three-fourth higher relapse rate. It was 15 percent.

So it was a drastic difference between the adherence kids and the nonadherent kids. And what was even more striking for me is that about 40 percent of the kids were nonadherent. So we are

Page 207

here in front of a life-threatening disease where we have a medication that can save those children, and 40 percent of them didn't reach a level of adherence what was enough to avoid relapse.

That brings me to the point that we have a major adherence gap, and I think we don't have a right vision about it because, typically, when we are talking about clinical trials -- except today at this meeting -- but usually when are talking about clinical trials, especially among practitioners, they say, oh, on trials, everything is perfect. So we have an idea that they think we are measuring treatment efficacy, while in practice we are measuring treatment effectiveness.

But given the data that we have collected, adherence data that we have collected in clinical trials, my view is that we are measuring something in between because we have suboptimal adherence in clinical trials, but we still do better than in practice because we select better-off patients and we do better patient follow-up.

But that's a very important point because

Page 208

that means that we end up with an estimate in clinical trials that doesn't answer the FDA question, which is, what's the efficacy of the treatment, and which doesn't answer the payer's question, which wants to know what's the effectiveness of treatment.

So that means that at the end, when we conduct clinical trials, we don't answer questions appropriately. We don't know the efficacy and we don't know the effectiveness. We are in between.

That often leads in drug development to the failure of the clinical trials to poor estimation of efficacy, in my view; also, inappropriate regimens. I will come to that point, but I think very often we go to a too high dose. And it's becoming important, I think.

The topic of adherence, I was very pleased to see in the enrichment guidance, the draft guidance from the FDA, it's mentioned. And I think it's very important that in the future we take into consideration this very important aspect, which is medication adherence, because when we think about

Page 209

the sources of variability in drug response, how much attention we require on the manufacturing level, at the prescribing/dispensing, it's a lot.

When we see the work in drug development that is done to study pharmacokinetics and pharmacodynamics in the last 20 years, it has been a drastic improvement. When you look at how many studies fail because of pharmacokinetics, it's a drastic reduction because it has been a topic of study in drug development that has taken a lot of attention.

But it's amazing that nobody takes account of adherence, which is a major source of variability. Drugs don't work in patients who do not take them. Remember the first sentence.

So now I will discuss a little bit that impact from adherence, pharmacokinetics, pharmacodynamics. And when we think about what are the consequences of medication nonadherence, it's clear that drugs don't work in patients who do not initiate them. It's clear that drugs stop working in patients who discontinue them. But they key

Page 210

point here is also drugs work partially or even create harm in patients who implement it sporadically.

Here I would like to introduce the notion of drug forgiveness. That's something that we don't very well know, but we should better study the forgiveness of each treatment.

What do I mean by that? Is that when we prescribe a treatment, this is the typical pharmacokinetic profile. So we expect that when we prescribe a treatment, after a few days of drug intake, the patient reaches a steady state and maintains a steady state over time, and we hope that that drug exposure is within the therapeutic index, the therapeutic window. That means it's high enough to have effectiveness ,and it's low enough to avoid toxicities.

So now if we simulate here three missed doses, you can see that the treatment will still be efficacious for about 24 hours. So we can say that the forgiveness of this treatment, this hypothetic treatment, is 24 hours.

Page 211

Forgiveness is bidirectional. If the patient takes an extra dose half day, you can see there will be a peak, but it's also forgiving for that peak as well. And in the future, we need to better understand the forgiveness of the different treatment, which we don't do today.

Because in practice, when we prescribe a treatment -- here, for example, once a day -- we think the patient reaches a steady state after a few days and maintains a steady state over time. That's all the picture that we have in mind when we saw this patient is on treatment -- you agree, this is on treatment -- while in reality, the patient missed a dose, take an extra dose to compensate the missed dose the day before, takes a little drug holiday, and there is much more variability in that process that can eventually lead to toxicities or periodic loss of effectiveness. And when we think HCV or HIV, there is also emergence of drug resistance.

So there is much more variability that we think in the process, and we need to better

Page 212

understand what's happening in clinical trials especially.

This figure shows us also the ability from electronic monitoring to project pharmacokinetic profile, and that's very interesting. I show you two examples here. In a therapeutic area, when people think there is an adherence issue, the first reflex is to say, "Let's measure concentrations." That's happening, for example, at the moment with the NOACs, the new anticoagulants. There is an adherence issue, and the first reaction is, let's measure concentrations.

I show you here two patients for which we did therapeutic drug monitoring. So that means that at day 21, we did intensive pharmacokinetics. Patients were hospitalized here also. And then we collected one, two, three, four, one, two three, four expected trough samples. So we are asking the patients to come at trough just before the next dose.

What do you decide about this patient? This is a trough. This is a trough, this is a trough,

Page 213

this is a trough. Too high? Too low? Do you decrease the dose? Increase the dose? You don't know.

When you look at what's happening really based on electronic monitoring, you see that he was catched [sic]. He was not a trough. He was a trough. He was catched when he missed a dose. So without the electronic monitoring, it's impossible to make sense of those data.

Here do you think this patient is controlled? Probably yes. It looks like he is pretty good, while in reality he missed quite a lot of doses, and the variability in that profile was at high risk for losing effectiveness, but also for emergence of drug resistance in HIV. So that's why it's very important to better have this dynamic of drug exposure and to better understand that to make sense of all clinical trials.

This is the last example I wanted to show you. We were involved in a dose-ranging study, three groups. It was a cardiovascular medication. It was a twice-a-day dosing regimen. And this is

Page 214

cumulative dosing.

If you look, it was 90 days in the study, so the patient who is taking the full dose twice a day, he would end up at 180 doses taken here. He will be perfectly aligned with the upper green line. And every line, every orange line, is here a patient.

When you look what's happening in the placebo, there are deviations in adherence. But if you look at all those horizontal lines, patients ending horizontal line, were patients discontinuing treatment. So the major problem we had in the placebo group was discontinuation of treatment, nonpersistence.

When we looked at the 7.5 milligram, there was no discontinuation. All the patients persisted to the end, but there was a very strong trend to dose much lower than the full dose. So that means that in the 7.5 milligram group, the patients implement it in a way that they auto-adjusted the dose, but they persisted with treatment up to the end; while in the placebo group, they stopped.

Page 215

They stopped. A lot of patients stopped.

How can you make a decision about which is the appropriate dose if you don't know those adherence pattern? Because the highest dose is auto-adjusting to a lower dose. It's very difficult to pick the right dose if you don't know how the patients were really exposed.

So in my view, we do today a lot of adherence on informed clinical development, and it's very difficult to find what's the right dose to balance efficacy and safety because when we are in phase 1, we have good idea of formulations, we have good, excellent idea of PK/PD, but adherence is a big unknown. And the trend is to go to the highest safe dose to compensate for diluted efficacy, but there are unexpected adverse effects.

Those are the two reasons. Finding the right dose is the number one reason why treatment fail or are delayed at approval because it's very difficult to pick the right dose when you don't know what's the exposure.

That brings me to the slide, saying this

Page 216

Struthian approach is no longer an option because Struthian is the Latin name for ostrich, because in the past, we were dealing with treatments which has a very broad therapeutic window. If we think about statins, hypertension, the therapeutic window was very broad. And we could pick the high dose to build a lot of forgiveness in those treatments so that implementation was less important.

But when we look at the pipeline of the pharmaceutical industry today, we are talking more and more about treatments that have a very narrow therapeutic index. I'm thinking about oncology treatment. I'm thinking about the new anticoagulants. I'm thinking about MS treatments. The therapeutic windows are much more narrow today in those niche markets. And in that situation, it's key to identify what's the best dosing regimen, what's the best dose, and we cannot ignore mitigation adherence any more.

It's no more acceptable, it's no more ethical, to jack up the dose to a level to cover a population where half of the patients are not

Page 217

taking the medications prescribed. And then you end up with 40 percent of the patients on oncology treatment with serious adverse effects. This is no more acceptable.

So we need to manage adherence in the future. And in the past, when the patient was not taking his medication, it was always said, it's his problem. We prescribed the best medication for this patient; it's his problem.

But the cost associated to nonadherence becomes so important, so gigantic, that we need to take adherence into consideration. And it's an entire system. It's the patient, it's the family, it's the provider, the community, and the healthcare system.

But the point is that we are not improving adherence to improve adherence. We are improving adherence, and the objective is to achieve the best use by patients of appropriately prescribed medicine in order to maximize the potential for benefits and to minimize the risk for harm.

So we don't want to achieve 100 percent

Page 218

adherence in everybody. The objective is really to maximize benefits and minimize harm. We have to keep that in mind.

What we realized in doing a lot of clinical trials with electronic monitoring is that when the investigators started to see those data, they said to us, I need to have those data because I will manage my patients differently.

This is the first study we did in Belgium in about 400 patients, and what we did, we provided to the pharmacist -- we collected data for three months, dosing history data for three months, among those 400 patients, and we provided this dose data to a pharmacist, who could discuss the data with the patient.

You can see in this individual example that the behavior before the discussion and after discussion drastically changed. This is also reflected in the study globally; we had a 15 percent improvement in adherence.

The idea is here, really, this one. I show you four patients, and all four patients took

Page 219

exactly 75 percent of their prescribed doses during a three-month period. Look at the first patient. He has a problem in the evening. The second patient has a problem in the morning and in the evening. The third patient, when he missed doses, those are consecutive missed doses, which are the drug holidays. And the last patient, he implements perfectly but stops too early.

So those four patients have taken exactly the same number of tablets. You can imagine that the clinical consequences of those four will be totally different. But now also if you want to build an intervention, imagine that that patient show up at the investigator's site, at that visit.

You can directly focus your discussion and say, what's happening in the evening? Can we fix your habit in the evening? Could we fix and improve your medication intake in the evening? While this patient, he clearly has a barrier. He stopped for a given reason, and you can immediately discuss that reason with the patient.

So having those data, as the previous

Page 220

speaker mentioned the control process, I think if you have those measures, you can really intervene and build an intervention. Currently, there are about 20 studies that have done that, and the average improvement in adherence by feeding back those data in clinical trials is about 20 percent.

So what I say here is that the measurement is one aspect. Education to increase knowledge is key. Motivations to increase self-efficacy is very important. But by themselves, is not enough. We need a good measurement of adherence to increase patients' awareness.

You have to realize that most patients, when you show them the data, they just say, "Wow. Is this me? I never realize I miss so many doses." Because most patients, if you ask them, they say, oh, it happens. I miss here and there a dose. But when you monitor them, you realize it's 20 percent, and increasing the awareness is very important to understand medication adherence.

So what are the measures that we can use to summarize them? I think that direct methods like

Page 221

pharmacokinetics, pharmacodynamics, while a lot of people think that's the way to go further to measure adherence, the problem is that the sampling is way too sparse to measure implementation, and it's subject to white-coat adherence. People just take a dose before showing up at their scheduled visit, and it bias the estimation of persistence.

Self-report is affected by desirability bias, but also a lot by recall bias. You cannot ask a patient to remember the dose he forgot last month, you know? That doesn't work, by definition.

Pill counting has been shown to be censored by the patients and only gives an aggregate summary of adherence.

In the medical practice, I think it's starting to take up, but the use of electronic prescription databases and electronic refill databases gives us a very good estimate of the patients who don't initiate and the patients who stop taking medication. And it's underused today, but it will be more and more used, I'm sure, in the future so that we have good estimates, we can pick

Page 222

the patients who don't start, and we can pick the patients who stop.

But for treatments that require a precise implementation, electronic monitoring will be used. And for me in clinical trials, electronic monitoring gives you really a good idea of adherence and of drug exposure in the trials. And it's such a big source of variability.

When I see people doing pharmacokinetic/pharmacodynamic models and searching for weight effect, sex effect, those are minor, tiny effects. And then when you do adherence-adjusted analysis, 55 percent of the residual volumes is explained because it has a major effect, not taking the medication. And it's not an absorption problem. It's not an absorption problem; it's the fact that they don't take it.

So about electronic monitoring, there are several methods, and most of my research -- in fact, almost [sic] of my research -- has been based on detecting package entry. So the idea is that the chip is put in the package, and we detect when

Page 223

the package is opened.

So the first question I will get after the talk is, yes, but it doesn't prove it's ingested. And that's true. However, we did pharmacokinetic studies, a lot of pharmacokinetic studies, and we have a very strong relationship. We have less than 2 percent discrepancies between observed pharmacokinetics and predicted pharmacokinetics from the monitoring, electronic monitoring.

So in my view, it's a very reliable system because you have to cheat every day if you want to cheat the system, you know? You have to use it every day.

Recently, and it has been on the news in most of the countries, even in Europe, the system with the SmartPill. The idea is to put a pill in the chip so that every time you swallow the pill, and the chip for sure, you have to wear a patch, and the patch will detect the signal and send a signal.

So it's in theory a very nice system. But in reality, it's very intrusive, patients'

Page 224

acceptance is not very high, it's technically complex, there may be some safety issues, and the reliability is not that high. I just saw a paper and say, oh, it's 5 percent discrepancy between ingesting the pill and detecting the pill, and they are also affected with people taking PPIs, for example, because they need the acid in the stomach.

So the idea is nice, but my view is it's really get a lot of efforts -- difficulties, safety, technical aspects, burden to the patients to wear a patch, too little information additional to package entry.

Then technically, electronic diaries, or here also a new approach, which is taking a picture at the moment that you swallow the tablet and send that picture every day to the center to show that you have taken the medication, in my view, those systems have the problem that it adds a lot of burden to the patient.

The major issue is that we are trying to solve a problem, that is, patients don't do something. They don't take their medication. It's

Page 225

not in their habit. And what we are asking is, and, by the way, now you will take a picture. We are just adding burden and burden to the patients. So for me it doesn't work very well.

The two last systems have to be combined with SMS reminder, and it's burden and burden. The patients quit it very early, while package entry, in my view, it's a measure of pill in the hand. So when the pill is in the hand, it's a very good proxy that it will be swallowed. And the most deviations that you will observe is not opening the package, and when the package is not opened, it's not taken.

So we mentioned bibliometry. And just about the use of electronic monitoring, there are today 700 peer-reviewed publications, which have been cited this month 50,000 times, and the h-index is 112. So it's a very well-established, very well-documented method of measuring medication adherence, and I think it's really ripe to be used systemically in clinical trials.

There are major opportunities for adherence-

Page 226

informed clinical trials. It's about time savings because we have better-informed benefit/risk development decisions, shorter time to set the optimal dosing regimen; that's very important.

It's cost-saving because it's greater efficacy when the medication is taken, but also much lower variability due to variability in drug adherence, which increases the power and reduces the sample size. Fewer post-approval dose reduction. And at the end, the therapies will be improved because we have much more informative safety and more effective dosing regimens.

For example, if I take the NOACs today on the market, we have four NOACs on the market, they compete on adherence, and they have no adherence data. In my view, this is no more acceptable to be on the marketing side with competing nonadherence only because some are once a day, some are twice a day. There are issues there, and they have no data. And this is no more acceptable, in my view. Thank you.

(Applause.)

Page 227

MODERATOR: We have speakers from -- and -- there's one more talk. Is there? Sorry. Sorry, there's one more talk.

Sorry about that.

Presentation – Sarrit Kovacs

DR. KOVACS: Good afternoon. I'm Sarrit Kovacs, and I work as a clinical outcome assessments reviewer in the Center for Drug Evaluation and Research at the FDA. And I'll be providing the regulatory perspective on electronic data capture.

Disclaimer. So I'll be presenting my own views, which do not necessarily represent an official FDA position.

I'll be covering a few topics related to electronic data capture, EDC, including the types of clinical outcome assessments, or COAs, and modes of administration of COAs. I'll also speak about the regulations governing EDC, including FDA's guidance for industry, regulatory standards, and other published guidelines.

There are four types of COAs. One type is

Page 228

patient-reported outcome, or PRO measures, which is a direct report of symptoms felt or functioning experienced by the patient, for example, pain, nausea, or physical ability. PRO measures can be completed at home or in the clinic. There are also clinician-reported outcome measures, where clinical judgment or interpretation is needed.

There's interpretation of patient's observable signs, behaviors, or physical manifestations, such as evaluating a patient's motor functioning ability or assessing a skin rash. These measures are typically performed in the clinic.

An observer-reported outcome measure is a report by a parent, caregiver, or another nonclinical observer regarding observable behaviors displayed by a patient, for example, crying, vomiting, clutching the stomach. This type of measure can be completed in the home or in the clinic.

Finally, a performance outcome measure is based on a task or tasks performed by a patient

Page 229

according to instructions that are administered by a healthcare professional, for example, assessments of gait speed, memory recall, cognitive ability, and this type of measure is typically performed in the clinic.

Electronic data capture principles and considerations that I'm presenting today focus on COA data collection intended for primary or secondary endpoints in clinical trials. However, what I present may also apply to other types of data collection.

There are two main modes of administration of COAs, paper and electronic. Within the electronic mode of administration, there are a number of subtypes: IVRS, Web- or browser-based, and handheld computer devices such as a tablet, iPad, or personal device such as a smartphone.

There are many advantages of electronic over paper formats. I have listed some of the advantages. One advantage is that with electronic modes, there is no need to manually enter the data into an electronic database for data analysis,

Page 230

which might introduce human error. Similarly, direct transmission into an electronic database may reduce risk to the integrity of those data.

Electronic modes allow alarms to be set at regular intervals or incoming phone calls when using IVRS to minimize the risk of missing data and to increase the potential for greater patient compliance. Electronic modes also allow for time and date stamps to ensure patient compliance in that the data was indeed filled out when it was supposed to be completed.

This helps to avoid the chance that a patient may fill out all of their paper and pen daily diary entries spanning two weeks' worth of data in one sitting in the parking lot immediately before handing them in to the investigator.

Another example of an advantage of electronic modes over paper is that patients completing electronic diaries can record when they take their pain or rescue medication and transmit those data electronically in real time. This way sites can see which patients are not compliant, and

Page 231

would know to contact those patients to make sure that they're complying with taking their medications.

Electronic COAs, just like paper COAs, need documentation of development and validation for FDA review of evidence to support labeling claims. The FDA's PRO guidance describes good measurement principles for developing PROs. Some of these principles may be applicable to other types of COAs as well.

It is important to note that the PRO guidance provides an optimal approach to PRO development. However, flexibility and judgment are both necessary in order to meet the practical demands of drug development such as tight development timelines. In addition, the FDA encourages drug sponsors to engage in early and continued communication with the agency during instrument development and evaluation.

With regard to electronic COAs, additional documentation may be important for FDA to review such as design features, like skip patterns and

Page 232

forced response. For example, it may be recommended that sponsors include a "not applicable" choice to avoid inaccurate data when patients are forced to enter a response for every question, like was mentioned earlier by Sharon Hertz.

Additionally, it would be helpful if sponsors describe their plans for addressing missing data and analyses. Even when sponsors implement a forced response, patients can turn off their device, which may result in missing data.

Sponsors should include for agency review any device usability testing and results, as well as patient, investigator, and site training materials, and documentation related to migrating or reformatting an existing paper instrument to electronic format.

There are some device-specific regulatory issues that FDA reviews, such as comparability of the data obtained via different collection formats. For example, some studies include the option to either bring your own device, where patients use

Page 233

their own personal devices such as smartphones or tablets, or the option to use a site-provided device.

There are some differences in these formats that may be important to consider, such as when patients use their own device, it's assumed that they have it on them at all times, which may aid in patient compliance. This may not be the case with a site-provided device.

In addition, investigators must ensure that the device should be available to the entire enrolled population. If studies include only the BYOD option, this would likely exclude potential patients from enrolling or participating in the study.

Investigators must make sure that replacement devices are available in case of device failures or lost devices to minimize the risk of missing data. And investigators must also include data entry date and time stamp documentation for agency review.

There are also some data-related regulatory

Page 234

issues that are reviewed by the FDA. Sponsors and investigators must ensure that the FDA regulatory requirements are met for recordkeeping, maintenance, and access.

The sponsor responsibilities are independent of the method used to record data, in other words, paper or electronic. Sponsors should plan to establish appropriate system and security controls as well as cyber-security and system maintenance plans that address how to ensure data integrity during network attacks and system updates, software updates.

It is important that sponsors establish a database backup as well as take steps to avoid premature unplanned access to unblinded data. Many of these regulatory issues were mentioned earlier by the previous presenters this morning.

The investigator is responsible for maintaining direct control over the source data, providing access to the records that serve as the electronic source documentation for the purpose of an FDA inspection or verification of source data.

Page 235

Exclusive control over the source data by the sponsor must be avoided. The clinical trial protocol or another document should specify how the electronic COA source data will be maintained and how the investigator will meet the regulatory requirements.

Direct electronic COA data transmission from the electronic data collection device to the sponsor, clinical investigator, or third party must include an electronic audit trail that documents all changes to the data after it leaves the electronic data collection device.

There is FDA guidance for industry available pertaining to electronic data capture. First, FDA's PRO guidance describes specific concerns when using electronic instruments, details the sponsor and investigator responsibilities, and provides warnings regarding practices that sponsors should avoid.

The FDA's guidance for industry on computerized systems used in clinical investigations provides to sponsors, CROs, data

Page 236

management centers, clinical investigators, and IRBs recommendations regarding the use of computerized systems in clinical investigations. And this applies to records in electronic form that are used to create, modify, maintain, archive, retrieve, or transmit clinical data required to be maintained or submitted to the FDA.

The FDA'S guidance for industry on electronic source data in clinical investigations includes recommendations to sponsors, CROs, clinical investigators, and others, ensuring the reliability, quality, integrity, and traceability of data from electronic source to electronic regulatory submission. And these two latter guidance documents are intended to supplement one another.

FDA's Code of Federal Regulations Part 11 is related to electronic records and electronic signatures. And this includes the criteria under which the FDA considers electronic records and signatures to be trustworthy and reliable and generally equivalent to paper records.

Page 237

This regulation requires FDA-regulated sponsors to implement controls, system validations, audit trails, authority checks, electronic signatures, and device and system checks.

The FDA's Code of Federal Regulations Part 312 relates to investigational new drug or IND applications, and Part 812 relates to investigational device exemptions. These regulations apply equally to both paper and electronic records. They include the general and specific responsibilities for sponsors and investigators with regard to recordkeeping, maintenance, monitoring, and allowing FDA access to records for investigation.

Electronic COA data must also be compliant with International Conference on Harmonizations guideline for good clinical practice. Specifically, sponsors must ensure and document that the electronic data processing system conforms to the sponsor's established requirements for completeness, accuracy, reliability, and validation, and that is consistent with intended

Page 238

performance.

They must maintain the standard operating procedures, or SOPs, and ensure that the system permits data changes so that they are documented and maintain an audit or edit trail. Sponsors must maintain a security system as well as list of authorized individuals who may make the data changes and adequate backup of the data, and safeguard the blinding of those data during data entry and processing.

This slide and the next include the relevant references and links that I mentioned during the presentation. Thank you.

(Applause.)

MODERATOR: Thank you, Sarrit. I'm sorry for missing you.

DR. KOVACS: Thanks. Oh, no.

Q&A and Panel Discussion

MODERATOR: I was looking on an old agenda.

Are there any questions for Sarrit? If not, we'll go ahead and go to the moderated session.

I just want to thank each of the speakers

Page 239

for preparing some great talks. They covered a diversity of issues. We started with pain reporting, so just wanted to start out there and just ask a really quick question.

The problem of having patients who report their average pain being higher than their worst pain or lower than their least amount of pain goes to issues in numeracy. And we talked about this briefly, Mark, about whether or not it might be useful to have a numeracy screening instrument.

I was just curious about, well, one, whether or not you might want to articulate how you feel about that. But then, two, has there been much work done with VAS where the VAS corresponds for VPI? Does the average more often exceed the worst score on the VAS versus the numeric rating scales, or are there differences there? Or is anyone familiar with that?

DR. JENSEN: I'm not aware of any research that has compared the VAS versus the NRS with respect to that particular problem.

MODERATOR: Issue, yes.

Page 240

DR. JENSEN: In respect to the first issue, you know, the rates of people actually rating their worst pain as lower or higher than their least or worst pain or average pain are not that high. They do exist, but they aren't that high.

It seems to me that if you want to screen for that problem, it would be easier just to screen for that problem rather than assess how good somebody is at math before putting this patient on trial. That's my own sense.

MODERATOR: Right. John, please?

JOHN: With regards to all of the issues that were discussed, in particular the issue of measurement and enrolling patients who may be professional patients or others, I just wanted to be sure to ask and to make sure that the speakers agree that what we're talking about here is a potentially nondifferential problem, meaning that ideally, if the study is properly blinded, then patients would be randomized to be in either the placebo or the treatment group with equal probability, raising all kinds of issues about

Page 241

noise and reducing the ability to detect real change.
So I don't downplay the problem. But I just want to be sure that we're on the right page here, because I think with each of the issues that we're considering over these two days, we need to think about how it might affect the results.
I think the important part of this one is that if the study still shows a positive effect, that the issue of professional patients doesn't negate that finding. And I wondered if any of our speakers wanted to comment on that.
DR. DEVINE: I'll take the first stab at that. I think professional subjects in the study, at least for my discipline, there's certainly risk to them. But in terms of the risk to the integrity of the study, we're worried that too many professional subjects will limit our ability to actually answer the question.
But also, I think that you need to entertain the possibility that people will exaggerate a disease condition in order to get in. So in my

Page 242

field, they're telling me they're drinking 28 drinks a week for men to qualify for my studies, and once they're in, they're more truthful about their drinking.
So there is this tendency to start high and then low, showing the possibility -- now, let's assume it's evenly distributed. That could happen. But I don't like that potential trend for them to start high and then low and show the possibility of efficacy.
MALE SPEAKER: I would sort of agree with you in terms of if you had a significant effect, the presence of professional patients doesn't take that away. But more important than is the effect significant or not to me is the estimate of the true effect. And it messes that up.
So I love the idea of strategies for limiting the people that enroll who are professional, and if they're identified later, it seems to me that a trial ought to have permission to remove those from analyses if they're later identified.

Page 243

JOHN: I'm not sure I agree with you on that particular issue. But I think the point you're making makes huge amounts of sense. We need to be a little bit careful about how far down that road we want to go.
No study is ever going to be perfect. There are going to be some patients that sneak through the process. There's no question about people complaining more about symptoms to get into trials. It's a large part of the regression to the mean process or the natural history of the disease process.
But my point was that we should try to eliminate that as much as possible, but not go head over heels with finding only the three perfect patients in the world to enroll in our study, and that part of the reason for that is that it's not going to give us a positive result when a positive result does not exist. It may well lead to a negative result, which I think is an issue with regards to efficacy.
DR. DWORKIN: I generally agree with that,

Page 244

John. But one concern is if you have a percentage, say, of 20, 25 percent of the patients in the trial are either duplicate within the trial or across multiple trials, they're not taking the drug.
So that means your safety signal is going to be some underestimate if a quarter of the patients in the trial aren't taking the drug, are nonadherent to medication. So that's an issue with the professional patients because they're smart enough to realize that drugs have risks so they don't take the drug.
The other thing. It's not professional patient. Remember *Jeri from East Palatka, Florida, with relapsing-remitting MS. She was blogging with patients in a trial to basically encourage everybody to unblind themselves. And that could lead to a false positive result.
In a placebo-controlled trial, if patients are blogging about the side effects of active, and the absence of side effects with placebo, and a percentage of patients in the trial become unblinded because of that, you can get a real false

Page 245

positive result because of the patient unblinding.
So I think we need to keep in mind the professional patients who are duplicating and who are really some kind of misconduct, and then the Jeris from East Palatka who are unblinding themselves.
JOHN: The unblinding issue is separate. I completely agree.
MODERATOR: We're going to go to the back. I'm sorry, I don't see your name. Could you --
DR. RICE: Andrew. It's because I've turned it over. Andrew Rice from London.
I'm asking this question, really, because I don't quite understand something coming from Europe, and that's this concept of professional patients. And I was talking to Philip over lunch about it, and we shared some common experiences.
I wonder if someone can just, for the sake of the few Europeans here, explain to us a little bit more about them. I'll tell you how our conversation goes.
Surprisingly, there are no European statutes

Page 246

or regulations, and UK ones, for inducements to participating in clinical trials. Each ethics committee can decide its own. But by and large, you're allowed to refund -- and I'm not talking about phase 1 -- reasonable expenses.
To us, reasonable expenses usually means a fairly modest public transport reimbursement. Many of our patients -- Philip and I both share this -- actually just refuse that altogether. They don't even claim.
Generally, we're not allowed to reimburse people for lost time at work, and I wonder if that's the difference here, because I can't quite understand how people can earn these kind of monies if you have much the same regulations as we do. And it might be that reimbursement for work issue that's the difference.
Just a quick rider on that. We were also wondering whether, having listened to the statistical monitoring talks, whether monitoring of the expenses at each centers, and if you start to see blips with large expenses, implying large

Page 247

distances of travel, you might be able to detect patients playing the game. So we put it that way.
But we wonder if there's a big difference in reimbursement that's allowable between Europe and the U.S.
MALE SPEAKER: So, Andrew, I don't know if that's it. But there are two ways that patients can be reimbursed for travel. They could actually get reimbursed from like a taxicab or a bus receipt. But what's also often done is they get a certain amount per visit, and that's considered, for the purposes of the ethics committee, modest.
But that modest could be somewhere between 25 and $50. And if the visit involves QST, for example, it's going to be closer to $50 because it's uncomfortable and it's time-consuming. And if you think about the professional patient who's participating in three trials at the same time, 100 to $150 a week is a reasonable amount of money. And our economy isn't as good as the economies in Europe, so it might be that that's part of the difference, too, that $150 in the United States

Page 248

goes further than it does in Europe.
DR. RICE: The economy in Europe is not as good as you might imagine.
(Laughter.)
DR. RICE: But that is a clear difference, and it may be something worth pointing out. And actually, comparisons of the behaviors in patients in clinical trials in those two systems might be a worthwhile thing to suggest because, as far as I'm aware, and Philip can tell me I'm wrong, we don't reimburse patients for their time. So if they had QST, they wouldn't get reimbursed. They'd just get their travel.
MALE SPEAKER: If you had a two-hour visit where you did the DFNS QST protocol and a bunch of questionnaires, what would a typical patient in London get for that two-and-a-half-hour visit?
DR. RICE: Travel expenses. They may get a cup of tea if it's a good day.
(Laughter.)
MALE SPEAKER: Yes. We give no money at all for that. And our patients, as Andrew said, often,

Page 249

if we offer them money for their travel, they'll reject it.
MODERATOR: Wow.
DR. VRIJENS: But is there not also in the U.S. patients who participate in the trials, they get health coverage?
MALE SPEAKER: No.
DR. VRIJENS: While in Europe, we're all covered, which makes a big difference.
DR. CONAGHAN: I think the drug -- that's important for free drug. That is a difference with our health systems. And we do see people trying to get into the one or two studies where there's a new drug that the health system hasn't okayed. So I'd imagine in the U.S., that's quite a big driver.
MALE SPEAKER: Just a point of clarification. In the U.S., there's safety oversight in terms of healthcare. But there's not comprehensive medical care provided for other issues outside of the -- that's my understanding, at least. John?
JOHN: Patients still get healthcare. But

Page 250

certainly we advertise studies that say, come in and have a free EKG. Hopefully, that's all getting better. I'm sure you've followed the arguments in the U.S. about healthcare. But anyway, then hopefully that goes away.
I think the other issue, though, is maybe just a difference in philosophy. But there certainly are patients who don't have a lot to do with their time and will travel substantial distances to do this.
I guess there's a real question as to how important that is in some of our trials. I can understand that in certain kinds of trials, it would be very important. And we worry about this in trials with opioids, for example, where there are a number of patients who try to get into the opioid trial because they want to be on opioids. But I think otherwise not.
MODERATOR: Thank you. I think Mike Rowbotham, and then we'll go to Roy, and then we'll go to this side over here.
DR. ROWBOTHAM: Yes. My question is for the

Page 251

panel but also for the previous speaker as well, Dr. Kovacs. I haven't heard anything about using telemedicine or other techniques to try and do two things. One is to have more central study monitoring and verification that this is a real live subject in a way that's HIPAA-compliant or other kinds of biometrics that would allow you to pick up subjects that are in multiple studies at the same time, even across trial sites.
So I just wondered if we could get some comment on telemedicine and similar techniques because those are really coming much more into mainstream medicine now.
MODERATOR: You want to --
PANELIST: In our studies, even that are multisite, we like to have one center that does data collection. We train the interviewers, and we have a live person who calls up and does the interviews, and they're trained and such.
So it's not electronic, but it's not paper and pencil. It's actual phone calls. It seems to work really well from my perspective.

Page 252

DR. DWORKIN: At my medical school there are a bunch of movement disorder specialists, Mike, who are researching, so it's not really live yet, but researching doing, say, the majority of clinical trial visits using telemedicine.
Patients like this, of course, because it means they don't have to schlep from Buffalo to Rochester. They can do visit 3 in a kind of Skype, but HIPAA-compliant, link. And so these movement disorder specialists think this is the future because patients like it much better, and it's cost-effective, and a lot of the visits can be done that way rather than requiring the patient to come into clinic.
PANELIST: It reduces missing data, too.
DR. DWORKIN: That's right.
DR. VRIJENS: Also, when you monitor medication adherence using electronic monitoring, over the 20 years, we have detected fraud several times. A center where all the patients take their medication exactly at the same time, that was in the early days. We have had that.

Page 253

But also, very recently we had a study and we identified a German site, which was not a real clinic. It was a professional site. And very strangely, all the patients in that site were like the Swiss train I showed you, the perfect patient.

There were some deviations, but they were -- and then when we looked at the clinical data, because it was hypertension, it was 100 percent success rate in that site. But the difference in outcome was 1 millimeter of mercury. It was nothing.

Then combining the adherence data with the clinical outcome, we could determine it was a professional site. But it's only possible to detect when all -- it's really recruiting only professional patients in one site.

But we saw also, for example, sites where -- it was a Polish site in one study, and they prescribed a drug systemically, once a day, to all their patients. It was a twice-a-day medication. We detected that very early on. All the patients were on once a day. It was a twice-a-

Page 254

day medication. The site didn't understand the prescription.

MODERATOR: Roy was next, and then we're going to go to this side over here.

DR. FREEMAN: This is a question to Eric. We can quibble about the effect of fabrication on drug effect and on adverse event profile. But no matter what, obviously your study is one of the most disturbing studies that we've seen in a long time. And you, rightly said, could be even worse. These are duplicitous people. These are fabricators. It's likely to be worse.

So let me maybe give another view, and that is we're used to, in clinical trials, patients coming to the trial saying, you know that nice Dr. Dworkin, he's so enthusiastic about this drug. Let me just make him feel good and tell him that it's working.

What about the equivalent in your study? That nice Dr. Devine, he really seems to want patients to be evil and fabricate and be duplicitous. Let me give him a nice positive study

Page 255

so he can get promoted.

DR. DEVINE: Perhaps I created a demand characteristic in my study, and elicited fraud.

(Laughter.)

DR. DEVINE: I suppose I can't rule that out as a possibility. But I have seen plenty of examples of fraud.

DR. FREEMAN: And I have, too. And I'll tell you some stories, too. I've seen it in non-drug studies. But I just want to give this perspective.

DR. DEVINE: Yes. When I did the study, something I didn't include in my talk, and maybe not even in the paper, is I used a slight bit of deception to lure them into telling me that they are using fraud.

The main purpose of the study and the consent was to evaluate rates of reimbursement and the discrepancy between what subjects think they should be reimbursed and what they're actually reimbursed.

(Laughter.)

Page 256

DR. DEVINE: So I started with those questions. And I actually had a publication from that plan that was looking at whether professional subjects are likely to fall into accepting risk that they shouldn't accept, basically undue inducement.

So there was a plan for that. But a little bit of a lure to get them thinking that maybe I'm on their side, and some normalizing around concealing behavior, before I asked them about the concealing behavior.

DR. FREEMAN: Then duplicitous on both sides of the table.

(Laughter.)

DR. DEVINE: Well, it was approved by the BUMC IRB, and the panel that it was approved by, I was not sitting on that panel. I recused myself.

MODERATOR: Go ahead.

DR. UPMALIS: Hi. I just wanted to --

MODERATOR: Could you say your name? I'm sorry.

DR. UPMALIS: It's David Upmalis from

Page 257

Janssen. But I think that the fraud is one issue, but I think there's a bigger issue that's a dynamic where you have a chronic illness, professional trial sites that are encouraged to recruit quickly, and oftentimes say, well, we have this huge database of suitable subjects and that sort of thing.

I really wonder if there isn't an issue where you get patients who can qualify for study after study, learn how to do that, and they come not out of any ill intent but out of good intent because they do like the coordinator, they like the visit they have, and everything else. And they're all such nice people.

It affects compliance. It affects continuation. They are encouraged to continue the study, so that you're getting a distorted picture. And these patients who are chronic, and I've had this experience not in a pain trial but in another trial, have all their medications that were prescribed to them still at home. And if they really want to stay in the study but they don't

Page 258

like what's happening to them, they can resume taking them, and more often than not, we don't detect that.

DR. DEVINE: I don't believe that we should recycle study subjects.

DR. UPMALIS: I agree.

DR. DEVINE: And actually, my sponsor is on the same page. It's a typical exclusion criteria for our trials. The past seven years of clinical trial enrollment for alcohol problems is an exclusion for participation in the study.

DR. UPMALIS: And in some trials I've done recently, I've limited it to two trials, and that's it. But I think that there's something there that needs to be done or considered.

DR. RAUCK: Yes. Richard Rauck, a clinician and academician at Wake Forest. I'll play a little devil's advocate. I think this is something we ought to be able to get out of some of our trial data. You guys did a great job with the pregabalin/gabapentin studies. And maybe alcohol is a little different than pain because chronic

Page 259

pain tends to be stable.

Recently, I was trying to think of a series of trials we've done. So there are all these ROOD trials, these rapid-onset opioid trials. They all came along at the same time. Right? And you had to have a cancer diagnosis.

So we found, what we thought -- because they were out of our practice, reliable patients. And it goes a little bit to your training because if they do re-enroll later into a very similar trial, those patients now at least understand the trial, how to record data. If they're good patients, I'm not so sure they don't give you better data and at the end of the day can look at a treatment effect versus placebo more honestly, more effectively.

So I'm a little concerned. Clearly, if patients are concealing or fabricating data, nobody wants those if they're doing that kind of behavior. But is that synonymous with saying an experienced research patient who goes through multiple trials is less reliable in the data they give? I don't know. We ought to be able to get at that, and we

Page 260

ought to try and find out.

It may be easy to know in the trials we've all done, right, who is a first time in versus somebody who's a repeat come in. And just like you guys did with the gabapentin and pregabalin trials, it would be great to know. Are they more reliable, less reliable, in differentiating a treatment effect or effect versus placebo? I don't know.

Maybe somebody has done that. Nat, have you done that?

DR. KATZ: Not yet. But I just wanted to add my voice to yours. I totally agree with you that if you go to experienced pain research centers -- at least I know a lot of people who do this -- the same patients can be great, one study after another. They understand what to expect. They understand what's required of them. They have a lot of experience monitoring their pain intensity. And you know from the last study that they showed up the whole time. They stayed through.

So we actually have a practice of

Page 261

encouraging our recruitment people to find people who've been good in previous studies. Obviously, as long as they -- also, we know them. We've got their medical records. And so I actually think that that's a good practice, and I think we're making a mistake by mixing the fraudulent or deceptive patient with the experienced patient, lumping them all together as the professional patients. I categorically disagree with that.

DR. DEVINE: I disagree. I don't see the professional subject synonymous with the repeat subject. But there are issues with re-enrolling a subject, at least in my discipline, where there's -- we fight very hard against the nonspecific treatment effects.

Assessment has a therapeutic intervention for alcoholism. Relationship-building has a therapeutic effect. And so people coming and getting to know us well and coming back actually introduces some error variance where they're actually benefiting from a psychosocial intervention, which is the relationship we form

Page 262

with them over years.

They've also had exposure to the behavioral platform that we use in conjunction with our medication. Everybody gets it. So there is something different about re-enrolling people that might get that extra bit of treatment.

DR. KATZ: I don't doubt that there are certain circumstances in which re-enrolling patients is bad and should be discouraged. But I just don't think we should paint it all with the same brush and say that repeat patients are always by definition harmful to our objectives in clinical trials.

PANELIST: So clearly, Rich, we can get the data from your site, whether this patient has been in a clinical trial at your site previously. Do you capture at the beginning of a trial how many trials they've been in, say, at any site in the past two or three years?

DR. RAUCK: That's a great question. We don't historically. So you're right. It would be hard, I guess, to go back. It's something we

Page 263

maybe, moving forward, should look at. I mean, I agree completely with Eric. There is some bias or something in the re-enrolled patient. There's no question that's a subset.

The issue to me is, does it give us better data, though, as far as does it separate real effect from placebo? Which is what we want. The industry more than anything just wants to know, is there a real treatment effect? And maybe it's better, maybe it's worse. I don't know.

I agree with you that it's an inherent subpopulation. I just think it would be great. Industry would probably love to know whether it helps give you a better definition of real treatment effect, I would think.

MODERATOR: We're going to go to Ian and then to the back there, and then back to Ajay. So Ian, go ahead.

DR. GILRON: I have two questions for Bernard. Thank you for an exciting talk. And I certainly think it's an important issue. I just wonder if there's any data to suggest that

Page 264

medication adherence is less of a problem in conditions associated with symptomatic therapy.

So someone gets up in the morning and says, I have a higher risk cardiovascular events and I need to take my statin, or I'm having pain and I need to take that. So that's one question.

The second question is, and you referred to this but I'm not sure, do we need to be less worried about medication adherence in phase 3 trials? And is most of the emphasis of what you were talking about more in earlier phase trials?

DR. VRIJENS: Let's start with the second question. I think we need to address adherence throughout the drug life cycle, and at the beginning. So very early on, we need to maximize exposure. And I would say it would be even more biased than today in trials, and force adherence as soon as possible, the full efficacy of treatment.

When we move to the next phase, I would say, in phase 3, for example, we need to understand much better what's the implication of nonadherence. And that's where you will start to learn about the

Page 265

forgiveness of the treatment in real population and so on. But it's very important to measure that and to see how forgiving is the treatment.

Then when you go commercial, then you know exactly what are the strengths and the weaknesses of your treatment. And then you can apply the appropriate methods for commercial. But it has to be taken throughout the whole life cycle. That's my view.

Now, for treatment with the feedback, we have done some pain studies. We have done some migraine prevention studies. Where there is very strong feedback also is PPI studies for gastric reflux. What we see is that there is a very strong selective nonadherence. That means patients feel better. They stopped. They quit for a while. Then they feel worse. They take. And you see all types of behavior there, especially in one of the PPI studies, which was prescribed on demand.

You see all the behavior there. You see the ones who continue taking it perfectly every day, and you see the ones who take it really as a

Page 266

feedback. You see the ones who skip a day systemically. You see all over the picture.

So in my view is the stronger the feedback, the stronger is the selective nonadherence.

MODERATOR: Is there a question in the back back here? I saw a hand earlier.

DR. JUGE: Yes. I had, I guess, two points and maybe a question on something --

MODERATOR: Could you say your name? I'm sorry. I really can't see your plaque.

DR. JUGE: Dean.

MODERATOR: Dean. Sorry.

DR. JUGE: That's okay.

MODERATOR: For the transcript, we'd like everyone to --

DR. JUGE: Oh, there you go. My name's Dean Juge. Two points, and they go back, one of them, to the variability in the scores and looking at patients that -- if the variability or the ability to collect the data ahead of time and say, well, these patients might be excluded from the study because they're more variable, my suggestion is

Page 267

that that might be a subpopulation to look at as an outlier group, not necessarily in the data collection for the primary endpoint for drug approval, but consider them your first set of phase 4 data.

Because what happens is you get drug approval on those that fit tightly, and then when the drug's on the open market, you have a lot of issues not seen that end up with problems later. And immediately, a lot of times the drugs get recalled when they could have foreseen programs and trained around.

There may have been a narrow therapeutic index. It may have been an adverse event, not really noted but something that's manageable. But it's just yanked because the risk of bringing it out is worse than trying to deal with the problem ahead of time.

So if there's a way to collect that data set, to me that's always the outlier data set that might be of value later to look at as what may be these issues that may come out of that group. Or

Page 268

for the standard practitioner out there, my concept here is how do you take this data from research to reality?

We do all these studies and research, but we also want the people to use this information and data to treat their patients routinely, and patient-reported outcomes or whatever to be a routine part of treatment later after the study's approved.

So that goes to my second point -- that's one group. But the second point is, you made the discussion about training or not and is it worth it. And from my perspective, it's worth it from two avenues.

One, the training is of value to show what's required of the patient, but if you study the differences in training, that's a huge value to payers because right now you have groups out there with people that you're dealing with. And if you can show that the training made a big difference in there, the payers might realize that a few extra minutes on a new patient might be worth an extra

Page 269

cost for the training or develop something that's worth the extra payment because it showed a difference in the trial when we did that with our patients. People need to understand that.

So if you have a study group that were going through the process, and then you had a routine training three months later and better development or whatever that showed that I'm constantly bringing this patient back in because they need more information or more whatever from these particular studies, that's something extra that, from a payer perspective in trying to get the patients in and getting them to understand this particular medication or this treatment or this whatever, might require something a little different than you've normally known. The data is there, and they can't just say, well, we're not going to cover that, we're not going to do whatever, and then those patients get dropped. So there's another reason to capture that type of data as well.

DR. DWORKIN: Speaking to the first point, I

Page 270

completely agree that in a phase 4 effectiveness trial, we're going to exclude a much smaller percentage of patients than phase 2 or phase 3, for exactly the reasons you said. I'd completely agree with that.

MODERATOR: So Ajay, and then we're going to go to the back corner there. Then we'll come back.

DR. WASAN: Thanks. This is Ajay Wasan. Just a comment on adherence, and I want to get some of the panel's thoughts.

So using the electronic chip with the package may be a gold standard. But a lot of us use electronic diaries daily in our studies, or once a week we email a Web link to someone to complete surveys. And I think one thing that Rob Edwards and I have found is that adding just one or two adherence questions like every day can become very helpful, such as, how many pills did you take yesterday? Or what time of day did you take your first dose today?

What I wonder is, and I want to get your thoughts on it, whether that's some degree of kind

Page 271

of a low-hanging fruit that we can add to the paper about recommendations for things that maybe could be done to better at least track adherence. It's not necessarily a gold standard, but it's using what we already use now, just a tiny modification. So just your thoughts on that would be good.

MODERATOR: Bernard, would you like to --

DR. VRIJENS: The use of diaries for me is -- the first thing is paper diaries. They are all filled just before coming at the visit. That's the first thing. Electronic diaries, they don't work in practice, according to our experience, because people tend not to use them.

They only worked when we start to have a reminder, you know? You have the diary, and then daily sent a reminder to fill that diary. Then you start to get some answer.

The problem there, in my view, is that when we are thinking about medication adherence, it's building a habit in the patients. And when you look at and you ask the patients what triggered the intake, they say, it's breakfast. It's dinner. I

Page 272

put my medication on my coffee machine, so when I take my first coffee, I see my medication.

When you look at the patients I showed you, most of them have about a three-hour average of taking their medication. So it's a three-hour window within which they take their medications.

So when you start to send a reminder to those patients to ask, did you take your medication -- because that's how a diary typically work; did you take your medication -- when do you send this text message, send this message, do you send it like an agenda, 10 minutes before, or do you send it 10 minutes after so you leave some time to the patient to do it? Or do you send it three hours after because we see that the average time, the average window, is three hours?

But then if you allow, for example, an hour, in most case, he will get that message while he's in the car and the medication's at home. Did you take it? Oh, yes. And so you get a lot of bias in those questions because it's always associated to time, and time is not a good trigger for adherence.

Page 273

So that's my point about diaries. Or you have to ask, yesterday what did you do? But you know most patients don't remember what they did yesterday.

A lot of patients -- it's amazing, because a lot of patients, they like to check they have executed the habit. How many patients, do they say, (gasp) "Did I take it today? I don't know."

DR. WASAN: I guess just a little -- in pain in general, I think all of us would say that we've found the electronic diaries to be very useful and a very high compliance rate. So I agree, yeah, there are some issues with adding adherence questions. But it doesn't necessary have as many pitfalls as you've describing. But anyways, interesting.

DR. DWORKIN: But Ajay, if you could afford it, so if you were a rich drug company or you had a big NIH grant, why wouldn't you have the electronic medication packaging? I can't think of a reason why you wouldn't implement that. I was quite --

DR. WASAN: Right. Exactly. That's why I

Page 274

said it's kind of a gold standard. And there's other similar things for --

DR. DWORKIN: So in a situation of not having the resources to do it, what are alternatives that aren't going to work as well that are cheap? But if you can do it, I can't think of a reason why you wouldn't.

MALE SPEAKER: I can. We've played with it here in a number of different situations, and the ones that you listed, which are -- it really grew up around HIV, when the initial HIV therapies had to be taken every five hours. Right? Not every four, not four times a day, but every five hours in order to reduce the viral load. Luckily, the treatment for HIV has gotten more consistent over time, and so it's less of a problem there.

But the issue with regards to pain, Bob, is that many of our patients are on more than one medication. And so a lot of our patients will say, I put them in a little box that I have. Now, you can rig those boxes to record whether they open them or not, and there are other ways around it.

Page 275

But the issue of just taking off the top of the bottle to take out a pill is not always as easy as is suggested because in fact, patients will tell me, I take out the two pills, I put them in my pocket, and I go to work, and I take one later. There are different ways of using them.

I'm not saying that the data's not useful. I completely agree that in the best of all worlds, it would make sense, perhaps, to do it. But at least in the studies where we're treating patients who have very significant symptoms, their issue is not about remembering the medication as much as it is about trying to get better and worrying about whether the medicines are going to work or not.

So if I had a lot of money, I would invest in some other things, perhaps, before I would invest in doing this.

MODERATOR: We've had some hands in the far back.

DR. CONAGHAN: Hi. I'm Philip Conaghan, Leeds. I guess back to Bernard on this same topic -- and to Bob and Dennis, thanks for a great,

Page 276

stimulating range of topics today because this is fabulous.

Bernard, how much data is there in the MSK field for MSK pain -- musculoskeletal pain -- using the electronic bottle capture data? I'm not aware of any studies, but maybe there are some. Because what I'd like to know is how much of an effect we can attribute to that adherence issue because we have great troubles with high placebo rates in all our studies. I think some of the chats here have really highlighted issues like pain variability and things.

Now, what if we can never change that? But adherence, we can. Adherence measurement, we can. So here we've got a measure -- I'd love to see a trial outcome that was efficacy by adherence as an endpoint for the trial. That would really mean something to me about efficacy, not effectiveness, but efficacy.

But have we got much data yet in --

DR. VRIJENS: Not that I know. We have data in rheumatoid arthritis, which is a different

Page 277

field. But no, I don't know that there are studies there.
DR. CONAGHAN: Because, Bob, I think this comes to the point you were just saying about putting it in trials. I would definitely put it in my trials now because I think, wow, maybe that's enough to start to separate out groups, allowing that we pay attention to the previous issues about getting rid of people we don't want in the trial. But this is a big -- we have to examine the question, even if it's not the answer.
DR. VRIJENS: Also, if you select the patients -- because you mentioned selecting the patients on training and viability. I have seen some depression studies where they were proposing and they were doing. They select the placebo responder in a running period, and then you -- because the placebo responders are the good adherers, and you just kick them out. And then you end up with worse adherers in the study.
(Laughter.)
DR. VRIJENS: So that's a big issue also in

Page 278

selecting based on placebo response.
MODERATOR: We've got Sharon Hertz and then Andrew Rice.
DR. HERTZ: Has there been any analysis of adherence in clinical studies and the relationship with clinical trial, which treatment arm people have been randomized to?
DR. VRIJENS: Which what?
DR. HERTZ: Treatment assignment. For instance, is there any correlation either between adherence and efficacy or, conversely, a lack of adherence and adverse events?
DR. VRIJENS: Yes. One of the first study that looked at this is the LRT, LRP LRT. It was a lipid-lowering study, a very big study. And they showed that the adherence to the placebo and the treatment were different. And that's where the placebo effect started to be discussed as well because they had a relationship between adherence to placebo and response to placebo.
So that started to trigger a lot of discussion among statisticians on how to analyze

Page 279

those data because there was this relationship between adherence to treatment and response, but there was also a relation between placebo.
That's because when you are adherent to the placebo, you are adherent to the other medications. I join your point because when you measure one medication, it's a proxy for the adherence to the other. And when you are adherent to the other, that's why the outcome was correlated as well.
So my point in the clinical trial on drug development, typically you are interested in one medication. And that's why monitoring that medication is important, in my view. And patients have to accept to use the package, and they have to accept not to use an organizer, to use that on top of it, which is something, if they accept to use a weekly organizer and this -- patients accept a lot of things in trials, so they have to accept this.
Now we are also involved in a lot of centers where they want to do this in practice. And in the elderly population, a lot of patients are on polymedications, and I think that the weekly

Page 280

organizer is a great, great tool to help adherence.
However, when there is an issue on adherence, I think putting one of the medications in a special box and monitoring that for a month or two months, building a habit on that one, focusing on that one, will help. And it has an effect on the other because I don't see in medical practice all patients -- even for a drug that has a very view therapeutic index, I don't see all patients forever on electronic monitoring. That will never happen.
However, when you have a very important medication -- I'm taking a NOAC, the new anticoagulants -- having them for one month to be sure he initiates well, a special starting program, that is something that may be useful in the future. But in trials, it's so important to know what's the drug exposure in trials. For me, it's a big help.
MODERATOR: Andrew, please.
DR. RICE: I wonder if we could just discuss briefly some other methods of monitoring adherence you briefly alluded to that we've had recent

Page 281

experience of. And again, like John, it comes from the HIV literature, where in the early days there was a lot of evidence of drug sharing in the very, very early trials. So that field is sensitized to the issue.

That's the use of unannounced random sampling of patients for plasma levels of the drug outside of the normal PK monitoring. We recently did that for an HIV neuropathy study, and despite all the normal test of adherence that you've referred to, there were a number of patients in the pregabalin group that had no detectable pregabalin in their blood, and even more disturbingly for a widely available drug, some patients in the placebo group who had pregabalin in their blood.

It's a very simple precaution, and I'm just very surprised we don't do it more often.

DR. VRIJENS: Yes. It works. These blood concentrations work if you do it -- you don't inform the patients when it will happen. It's at random. And there is an excellent study that is recent on the hypertension because they have

Page 282

developed a test -- it's a urine test -- that can detect about 90 percent of the medications used for hypertension.

It's not a measure. It's a dichotomic. It's yes/no, presence or absence of any medications for hypertension. And they did it without informing the patients, and they concluded that half of the resistant hypertensive patients were nonadherent, were not resistant. And they did it through a urine test like that.

But it only works if you don't plan it, and if it's acceptable in the population, that a patient walks in and you say, oh, I will measure you.

DR. RICE: No. I absolutely agree with you. It has to be done unannounced.

DR. VRIJENS: There are some studies they have also done, and they have compared it very successfully with electronic monitoring, is unplanned pill counting. So that's very frequent in California for homeless patients. They just walk in, they catch them, and they count their

Page 283

pills.

They have shown that the bias we see with pill count is not there when you do unplanned pill counting. But again, you have to show up in their home or where they live and catch their pills and count. It's very intrusive.

DR. RICE: But blood testing you can do on a visit where you were going to take normal blood for liver function monitoring or whatever, as long as it's not part of a pharmacokinetic part, where they come and take their drug at a certain time.

DR. VRIJENS: The problem is when you start to do that, people will know that you do that sometimes. And then it will be on the blood, and then --

(Laughter.)

DR. VRIJENS: But it works. It has worked in several studies.

MODERATOR: We're going to go back to David Hewitt.

DR. HEWITT: Yes. My question was, obviously it's good to find these things and then

Page 284

go back and retrain the patient and retrain the site. But from a statistical analysis point of view, you still are doing an intent-to-treat analysis.

Unless you're just going to use it for a sensitivity analysis at the end, it's kind of frustrating to know that at the end, you're using data that's bad, but there's not much you can do about it because you're still stuck with an intent-to-treat. So I was wondering if anybody wanted to speak to that.

MALE SPEAKER: We've talked about the usefulness of monitor and adherence, and we've talked about the usefulness of training for pain assessment. I don't think we've mentioned about whether, in trials, we should systemically train for adherence, provide these tips. And then maybe if in two weeks you see an upswing of lack of adherence, then that's the time to kick another refresher. But how many studies now systematically train for adherence?

DR. VRIJENS: Yes. My view was changed on

Page 285

that also because in the past, my first papers on adherence, I have always shown that patients do a wonderful experiment by taking less and more doses. They do a dose-ranging study. Because when you do a dose-ranging study, you change the dose. When they do a dose-ranging study, they keep the same dose but change the interval between dose.

So when you capture that information, it's very rich information because you can learn a lot from that experiment. And that was my view of using adherence data.

So I would say do the ITT analysis first. The ITT analysis is the base, it's clear. But then in addition, you can learn and you can have an estimate of a PK/PD model. You can have an estimate of full efficacy in addition to the intention-to-treat. That was my view.

But today, the FDA drug guidance, "Enrichment in Clinical Trials," they specifically mention that adherence could be used first for screening nonadherent patients in the run-in and kick them out. And secondly, the dose data should

Page 286

be used to maximize exposure and get full efficacy in the clinical trials as soon as possible.

So that's the view, and I think it makes a lot of sense because adherence-adjusted analysis could always be biased because it's a post-randomization variable. There could be bias there.

MODERATOR: Nat and then Jim Witter.

DR. KATZ: I had a quick question for Eric. I wonder if in the study that you did on the fraudulent patients, whether you ran across any stories of investigator collusion and those types of activities? Because obviously, it's not only the patient that has the financial incentive to get in the trial; it's also the investigator.

DR. DEVINE: Yes. That's an interesting question, which I did not investigate. But I would imagine there probably are investigators that look the other way. They know the subject has a disease, and they look the other way. So I might think about that for the next survey.

MODERATOR: Jim Witter?

DR. WITTER: Mark, one potential way to

Page 287

improve the sensitivity of the MPRS scale might be to go back to some things that Grace Lee worked with many years ago, where verbal anchors are added to the scale. This could be very simple, so zero to 3 could be marked with mild, 4 to 6 could be marked with moderate, and so forth.

So this would be, I think, a simple addition. And in my own experience one-to-one with patients, they are amenable to these kinds of suggestions, and they may use the scale in ways that don't confirm with the majority.

So they might say, well, my pain level is a 3, and at the same time they're saying their pain is really severe. But if you point out that it would be more useful if they used the broader range of the scale, they're okay with that. So this might be part of a training addition, and clarify the scale also to subjects.

DR. JENSEN: I don't know if that's included in your training, where you say, in general, people view this as -- that would be an interesting question. I don't know whether it would have an

Page 288

effect. But it could potentially improve sensitivity.

MODERATOR: Maybe, Nat, have you looked at this at all?

DR. KATZ: That's kind of a sad story. Actually, in our first version of our training program, we did have verbal anchors next to all the numbers. Mark, you may remember this.

It created a bit of a brawl; if you can imagine psychometricians brawling with each other, that's sort of what happened, where there was an armed camp that thought that it was a horrendous mistake we had made to include verbal anchors along the same side as a numerical rating scale.

So with neither data on each side, we just took it out to keep the peace. But I still think that it's an interesting idea, to see if we could train patients to do better.

MALE SPEAKER: It's amazing that that's an interesting idea at this point in the life cycle of pain science.

MALE SPEAKER: The U.S. military has

Page 289

implemented a pain scale that they are now testing, which has color, words, and numbers all on the same scale.

MALE SPEAKER: And faces, John.

MALE SPEAKER: I'm sorry, and faces. I'm sorry, I forgot to mention that.

MALE SPEAKER: Everything except the kitchen sink.

MALE SPEAKER: Yes. It'll be interesting to see how -- a lot of us commented about that particular scale, but they didn't think about it. They went ahead with it anyway. So they're going to be collecting a data. It will be interesting to see what happens with it.

MALE SPEAKER: So if empirically one added labels to a numerical pain rating scale, would that just be heresy to do that without the backup of validity data?

MODERATOR: I'd like to let Sharon Hertz comment next.

DR. HERTZ: I think you just answered the question. I think we'd like to know what that does

Page 290

to the performance characteristics, so you'd probably need to do some study before you wanted to rely on that.

I think everyone would want to know, not just from a regulatory perspective. But it will be interesting to know what that kind of -- does it really change it to categorical scale, or is it still a numerical rating scale? What does it mean? Because we know that 1 to 10 is not an even distribution, necessarily, of intensity. What does that do to it and how does that change the response?

So it would be important, I think, to have an idea of what you're doing. Just looking for concordant responses may not really be helpful if you don't actually know what you end up measuring.

MODERATOR: Jim Witter, would you like to --

DR. WITTER: Just another question again. I'm going to pick up on Philip's comment earlier. We in rheumatology have, over the years now, been blessed, I guess, with biologics, which are given,

Page 291

by necessity, by injection.

So what is the adherence, do we know, of injectables versus topicals versus per os medications, which is what I assume we've been talking about mostly here.

DR. VRIJENS: Yes. That's a big discussion because we have some areas, like oncology, moving from injectables to oral. We have the opposite in MS, for example, moving from injectable to oral, but also in RA, the new ones are oral. There are problems of adherence on both sides. So it's not because you go oral or it's not because you go injectable that you solve the adherence issue.

One of the first studies in leukemia, it was very clear, when you read the qualitative comments of the patients, it was like, if it was an important drug, they would inject me. So because it was oral, it was less important. So people perceived it very differently.

Then you have the oral issue of people don't want the needles, the fear of the needles. But typically, that disappears very fast. In medical

Page 292

practice also -- especially in Europe; I don't know in the States -- but all the injectables, MS, RA, there is a nurse to initiate the treatment for at least one year. They are very intensive the first weeks, and then they will follow the patient till one year.

The orals, they come in and they get a prescription, and it's oral. It's done. So that initiation by a nurse is a major, major difference in treatment initiation.

So we have monitored injectables, and we are starting a study last week in RA just with the injectables where we capture the time they throw away the needles. And we start to see some data in MS, but also in RA, that are really suboptimal in treatment.

Then those biologics, first of all, they are on top of other treatment, very often, and they're not by themselves. So it is also the adherence to the others. And the dosing regimens are pretty difficult, every other week; every other day in MS, every other day, every other week in RA. Those are

Page 293

really difficult to build a habit, and even every month.

The advantage with every month, you can send a reminder to the patient because it's only 12 reminders in a year, and it's not such a burden. But every month, it's a very difficult dosing regimen. It has been proven in osteoporosis where bisphosphonates, which have to be taken outside of food, they were very difficult to take every day.

So every week, it's easy to build a habit every week. Every Monday I do it. But when it starts to be every month, the results were not good at all because the first of the month is never the same day. It's very complex.

So we have to think about those. It's not because we go biologics or we go oral that we solve the issue. There are many, many issues, and we need to think about those.

MODERATOR: All right. Well, it's our coffee break now. We could keep going if people want to keep going, or people could come up and ask our panelists during the break. But thank you all.

Page 294

(Applause.)

(Whereupon, a brief recess was taken.)

MODERATOR: [In progress] -- on site selection and some of the details of central statistical monitoring. The discussion session will be held first thing tomorrow morning. So the plan is to end, I think, at about 5:00 today.

Our first speaker is going to be Rick Malamut, who's the vice president of global clinical development and head of the pain therapeutic area at Teva Pharmaceuticals. He's going to be speaking about site selection, training, and surveillance.

Rick?

Presentation – Richard Malamut

DR. MALAMUT: Hi. So again, thanks to Bob and Dennis for allowing me to come here. This slide is not depictive of what some of my employees think of me at the end of a day, but it is an example.

So I think this is probably well timed because as I've been listening through most of the

Page 295

day, not all, we spent a lot of time focusing on the study subject, the study patient, picking on them a little bit. And now I think we're going to switch gears a bit. And we want to talk more about the investigator, more about the sites, more about the monitors and what we could do better for that.

The other thing about this is I started to look for all the -- what I assure will be multiple research papers going into outcomes on whether it matters as to which site you pick and training. And we already heard from Mark about all of the studies he showed showing outcomes in monitoring. And I won't show all of them.

So a lot of this will come from internal, what we do at Teva, what may be standard across industry, so I'm very curious from my colleagues as to whether they follow these same guidelines, and then a lot about some of potential research questions: What should we do? Does training matter? Does surveillance actually matter and how can we actually prove that?

So I think this is familiar to all of you.

Page 296

A lot of this comes from the '96 ICH guidance, which is currently being revised or updated. And a lot of this has to do with safety and the rights of the patient and the person in the study. So it is about creating a high-quality research study that gives an appropriate result, but throughout it all, we ought to be mindful about the person involved in the study.

So I think a lot of this, many of you, if not most of you, are aware of as to what goes on internally when we put together a study, and so this is a slide that doesn't show everything that goes on. But I thought I would just show this, and then I would just highlight the two or three or four items that would relate to this topic.

As I started to look, I realized, wow, it's going to be at least half and maybe most of this. So the goal was to try to condense this down into 30 minutes. So I'll go through all of this quick, assuming that you-all know a lot about this, and then really to focus on some of the questions.

So first of all, site selection, so again,

Page 297

usually led by the operations group within the sponsor and is based on input from multiple functions, including commercial, though mainly, as you'll see in the third bullet point, based on the eventual plan as to where you might want to go assuming success. But certainly, the clinical group, the medical group who's out there interacting with the physicians who knows the individual countries.

Regulatory pathways that must be followed. Clinical drug supply matters between different countries and different regions.

The second bullet point can be applied to the site selection as well as country selection, but certainly, we want to know how well the country has done in prior studies of the same disease indication. What's the data look like? Has it been high quality?

We heard earlier from Sharon about a country that had some sites that maybe didn't deliver high-quality data, and that would go into our decision-making.

Page 298

How well have they done with the study requirements? We are interested in recruiting. Did the country seem to have enough of the disease indication with the right qualifications to bring into the study, and what is the disease indication prevalence? We wouldn't go to a country that doesn't recognize or doesn't have a high preponderance of the disease we're studying.

As I said, to get marketing authorization, which is the long-term goal, we may need to put a site or select a country. An example might be Russia, where you have to include a site in Russia in order to later get marketing in that country.

Then, of course, what's in the protocol? It may be differ based on the protocol. We talked about disease prevalence. Certain countries and regions are less accepting of including a placebo arm. What is the comparator when we're putting in an active comparator? In some countries, the comparator we choose may not be available.

What's the standard of care? What are study procedures? So all of this goes into selecting the

Page 299

country.

Now, when you get to the site, a lot of this is repeat, though with perhaps some additional nuance. So again, same ideas on the prior slide. We know that when you're investing in stocks, past performance is not a predictor of future success.

So in site selection, we don't follow that, and in fact, we actually rely heavily on how well sites have done in past studies when we're looking to choose them for the next study. And number one up there is quality. We are looking to make sure that the quality of the data that's been provided is actually high-quality data.

We would like the investigator to have some expertise in the disease area, particularly if it's a somewhat rare disease area where it needs a special level of expertise.

Unfortunately, logistics play a role. And if we find a very high-quality site with disease expertise and a large number of potential study patients, but we know the IRB at that site meets only every two months, and when we tried to

Page 300

contract with the site or that institution in the past, it's taken six months, regrettably, it gets very challenging to include that site as much as we may want to for the other reasons. So that does come into play.

Again, how well they followed study requirements. Did they not follow the protocol? What were the protocol violations? And then last is how well have they done in recruiting. And it's not just recruiting patients, it's recruiting high-quality patients, but yes, the numbers do matter in past studies.

If we have a less common or maybe more complex disease process, we often do, all the other factors aside, need to go to specialized sites that have those somewhat rarer patients. I guess arthromyalgia could be an example, where very few of these patients out there in the world and only a few sites may have access to them.

Or it may be that the protocol requires some sort of special skill. You need to be able to do nerve conductions or QST or some of the other

Page 301

things. Even recognizing what allodynia is and testing for allodynia. We all know. But some sites, is it better to try to train them and teach them or hopefully they actually know what it is ahead of time. I'd rather they know ahead of time.

So once we've made a preliminary list of sites, the sites themselves, look at the protocol, under confidentiality. They let us know their interest. Do they have the capacity? Are they involved in other studies? Do they think they can meet the recruitment goals? Do they have the bandwidth? Are they involved in four other studies, and they can't devote time to ours?

Once there's been a mutual agreement that we think they could participate, and they would like to, then there's a preselection site visit. So this is the [inaudible – mic fades] site selection, it's also the first step in training, and it's also the first step in surveillance, where our monitors and study personnel first get to go to the site, meet the site personnel, look at the facilities, and begin to assess whether they can actually do

Page 302

the study in the way that we want them to do it.

The last bullet point, if they've been previously inspected by FDA, we want to know how it went. What were the issues? Were they resolved? Were they resolvable? More ominous, we'll actually -- and we do this with every site. We compare a site that we think could, should be in our study with the FDA produced registries of sites that have been de-barred or have had some of these other restrictions placed on them. Warning letters, maybe not. We may investigate what that was about, but otherwise, we won't include sites that are on these lists.

So first break. And again, throughout this, since there isn't a lot of data to present to you, my conclusion overall was that we need data, and we need to know does this matter. Does selecting the proper site lead to a higher quality study with more reliable results, however that's mentioned.

So these are just some of my own questions. I don't have answers. I'm hoping some of you do. Opinions are welcome, and we'll get to some of this

Page 303

during the discussion tomorrow.

But what about the selection of an academic site versus a research site? The myth is that the academic site has the high-quality patients with the disease expertise but doesn't recruit so fast and takes a long time to get started. That's the myth.

The research site, the myth is, recruit incredibly fast. We'll bring in lots of patients, but they don't have the disease expertise, and a lot of those patients aren't the right patients.

So those are myths, probably like all generalizations, maybe not fully true, but what do we think? Does it matter in looking at the results of a study your proportion of academic sites versus research sites?

Next, what about those fast recruiting countries that we've heard about? There's a lure there. We know that we can go to certain regions, certain countries, and we'll recruit very quickly. Is the study quality truly not as acceptable as it might be in other areas?

Page 304

I talked about clinical study experience. So if we want study experience, how do we identify the next generation of high-quality sites and investigators if they haven't done studies? Is there something we can do? Is there some kind of training program? Is there some sort of identification process for a site or somewhere in the world where they might be interested but they haven't done it? And we can sort of nurture future high-quality sites.

What about speed and quantity? As an industry, we want to recruit quickly. We have our timelines. We have our metrics. So is that a negative factor in study data if you recruit quickly, or if a site has large number of study subjects, is that a good thing or a bad thing? There are some papers looking at the high enrollers with investigator enthusiasm leading to an increase in placebo rate.

So there isn't a lot out there. Do we know that a fast recruiting site necessarily has lower quality data?

Page 305

Training. So we do a lot of training, and everyone is familiar with training the sites at investigators meeting. What maybe not everyone knows is we spend a lot of time training vendors, and we'll get to a little about that with surveillance. And we also train our internal people.

There's an assumption that, oh, the sponsor knows everything that's going on with the study, and, in fact, that's not the case. Let me give you that insight.

So investigator meetings and other training meetings are to train everybody involved with the study. I think everyone here is familiar with an investigator meeting and what goes on there.

I want to look at the fourth bullet point where now there's a trend towards virtual investigator meetings with the idea that it will save money. I will tell you it may not save money. I'm looking out there and wondering if my brethren sponsors have seen the same thing. It's not so cheap to run a WebEx. And I wonder whether in

Page 306

trying to make things simpler, we're actually losing some of the advantages of a face-to-face investigator meeting. You may guess my bias.

Does it matter that you're meeting the investigators face to face? If an investigator is meeting the sponsor and the CRO, are they more likely to try to do things the right way? Are they paying attention? Are they even at the computer screen during the virtual investigator meeting? We don't always know. We think so. We put tricky test questions in to make sure that they're there, but are they actually doing it? Is it the coordinator?

Then by not having a face-to-face meeting, you need additional training afterwards to go through the vendor training. So it's a lot of indirect training. I wonder whether we had it right the first time, but want to know opinions.

Then after the investigator meeting, that's not the end. There's periodic training afterwards. There may be a need for an interim investigator meeting, and certainly, multiple teleconferences

Page 307

and WebExes between the medical monitor and the sites to discuss study issues.

I think everyone knows about the role of the principal investigator. Again, it's protecting the rights, safety, and welfare of subjects and certainly having control of the drugs and biologic products.

Look at that fourth bullet point. "Personally conducts or supervises the investigation." Well, that's the assumption. We know for those who have done studies, we have generally very well trained coordinators that we're joined at the hip with when we're running studies. So how much is being done by the coordinator? How much is the investigator involved, and does it matter if you've got an excellent coordinator?

What about the sub-investigators? We may vet the investigator as high quality, knowledgeable, but the person at the site may not be the investigator doing all of the procedures. They're listed, but do we have the same access for training and assessment? And then, of course, the

Page 308

investigator is responsible for all that happens. So hopefully, that's an enticement.

The site initiation visit comes generally after the investigator meeting and is another way to train the sites, and again, another way to begin some early surveillance because it's making sure that the site staff, including investigator, coordinator, and sub-investigator, truly understand the protocol. They understand the process of informed consent and so on.

Does the site have everything they need? Do they have dedicated locked drug rooms? Do they have a dedicated place to perform the study, or will it be done in the hallway because the exam rooms could be busy on a given day? And then to resolve the appropriate needs. The idea is not to be punitive, I'll say this again, but to actually help train and facilitate because if we've gone to this point where we believe the site can do a good job and they're appropriate, we want to try to help them to provide the data we need.

So informed consent, I won't go through most

Page 309

of this. I think everyone here is familiar with the concept of informed consent, but I want to look at that bottom box in which maybe as a way of improving data quality, this has been discussed before about trying to standardized the informed consent process.

We've heard talks from -- Nat and Neil Singla have both addressed this issue as to the value of trying to standardized informed consent to eliminate the overly enthusiastic investigator or to try to standardize for the potential study subject so that they're not coming into the study with expectations, negative or positive, that could impact the study. So that's something that I know is already being done and hopefully looking forward to further results of this.

So some questions about training. Again, I addressed some of these. What is the relative involvement of principal investigators and coordinators, and does it matter? And in fact, could it be better that a highly trained, invested nurse coordinator who's doing a lot of the

Page 310

involvement with the study subject, or do we really want the PIs or the sub-Is in the room or interacting more, taking more of the scales and so on?

We've already talked is there any difference in study conduct between a face-to-face versus a virtual investigator meeting. We bring people to investigator meetings, and we make this assumption that because we brought them somewhere and we gave them a full day of presentations on the protocol and study procedures that they've got it, and they're now ready to go out and do the study.

Or is that they have the investigator meeting, two months later, the first patient walks in, and everyone starts scrambling to find the protocol they were given to remind themselves, well, what are we supposed to do and did I have to do this now or is that later?

So is that enough? Do we need some kind of more formal certification? We don't have that really now.

What about overall certification? We know

Page 311

there's courses. We know you can be trained to do clinical studies. We don't require it. Should we? Should a site or an investigator or a coordinator be mandated to take some kind of certification? I'm betting that won't be a popular question, but I throw it out there.

Should we standardize the informed consent process? We're hearing again, as I mentioned, Nat and Neil might say yes, but we should talk more about that.

Study surveillance. We've already talked a bit about there's surveillance through the entire study, and there's some even surveillance activities that I haven't mentioned in here that come before the study even starts, and I'll show them here.

So we talked about the site initiation visit, the site activation where the site is visited and again, a surveillance before starting. And then while the study is ongoing, we have inter-monitoring visits where we send study monitors to the site, and we'll talk about that. And then the

Page 312

site audits, both internal from the sponsor but also in preparing for an external audit. And then the closeout visit, another means of study surveillance.

So inter-monitoring visits are mandated. They are necessary, but sponsors have quite a bit of flexibility. And I know in your pre-read, there were different types of central monitoring not mandated at all sites at all times, based on certain triggers. We wonder whether is that enough or should we standardize the monitoring, looking at the data. It is the monitor who is the primary point of contact between the site and the sponsor. They're our eyes. They're our ears.

We know we have inter-monitoring visits, but frequency is really based upon the individual site. What is the patient visit schedule at a given study? How is the site doing? A high enrolling site will probably -- not probably, will generate more monitoring visits. And again, helping the sites to prepare for internal quality assurance, internal audits, and then more formal inspection

Page 313

audits.
So again, the monitoring visit purpose is to make sure that the reported trial data is accurate; that it's complete; that where they were supposed to enter data, it's been entered; and it's verifiable from the source. We're going to come back to that in another way.
We want to make sure that the conduct of the trial is being performed following to proper requirements, not only based on following the protocol but also following GCP principles and regulatory requirements.
Are the drugs being stored properly? Are they accountable? Are they missing pills? When you walk in, are there pills all over the floor? Not a good sign if you walk in and you see that.
Are the safety events being reported? Are the monitors going in and hearing about a safety event that happened a month before and just didn't get around to being reported? That would be a bad thing.
Are the protocol violations actually being

Page 314

identified and reported? Again, monitors have gone to sites in some of our studies, and just like Sharon was expressing unhappiness that the FDA finds these things, sponsors are unhappy when we go in and we find things that haven't been reported to us.
Then again, it's not punitive. The idea here is not to say, oh, you're horrible, shame on you, you're out, you'll never work in this -- no. The idea is to try to train them or retrain them, make sure they're doing it properly, make sure they understand what they were supposed to do. Very often, they didn't quite understand. And then assisting the sites in resolving queries that come during the conduct of a study.
So the source documents, this will be one of my pet peeves is that -- so the source documents are the patient's medical file, the patient's medical records. And presumably, every study subject, every study patient has a medical file somewhere. Presumably, even healthy volunteers have a medical file somewhere. And so it has the

Page 315

history, the lab results, the medications and so on. The trick is getting to see it.
Again, if it's a site that where the study subjects, study patient is coming from the site, that's relatively easy because presumably, that PI has that chart in their office, or in their partner's office, or one floor down from the referring specialist.
But for some sites where a number of the study patients aren't coming from the site, they're coming from referrals, from advertising, at the research sites, maybe most, if not all, of the patients are coming from that route. So then we need to try to make sure that a patient who comes in and says, "Well, I had postherpetic neuralgia five years ago" really had postherpetic neuralgia five years ago. And how do you confirm that?
So there's a belief, my belief but also at Teva that, in fact, we need to -- in everybody, is that possible in everybody? But in every study subject, actually do a source document verification, look at that patient's medical files.

Page 316

Do they have a medical problem that they didn't tell us about? Are they on a med that would be contraindicated for the study? Do they really have the diagnosis we thought they had or maybe what they thought they had?
It doesn't have to fall into the fraudulent patient realm. It could be that the patient really didn't know what they had. They knew they had some kind of foot pain, and they had some kind of neuropathy. Maybe they didn't know what their neuropathy was or what the pain was from.
So we want to make sure and look at, as best we can, every single study patient's source document. And so our medical monitors do that. I don't know how standard this is. I'm curious to hear.
Vendors, we talked about oversight of the sites, but we also need to oversee the vendors. And many sponsors are more and more using external vendors to run studies, to be monitoring central laboratories, central EKGs laboratories, medical monitors. Sometimes the majority of the study is

Page 317

outsourced to a vendor, so the sponsor does need to have some oversight of what's going on.

As we outsource globally, we need greater scrutiny, not less. So just because you've outsourced it to a CRO, no matter how good they are and I'll say most CROs are really quite good, you still need to look. And sponsors still need to have a look and make sure that things are going well. And there have been some instances where the CRO didn't do everything they were supposed to do, and sponsors are responsible.

So again, Teva has an actual vendor management plan in which every study that uses external vendors, there's a vendor management plan organized between the sponsor and the vendor in which everyone knows what the roles are, everyone understands how the vendor's activities will be assessed.

These are just some of the things we may look at.

So data is important, and we've talked about collecting the data. There was an old statement I

Page 318

used to hear that if it's not recorded somewhere, it never happened. In a clinical study if it's not written down or entered into an eDiary, it didn't happen. So we have to make sure that the trial master file is complete and includes all the necessary documents.

Again, some of it has to do with site compliance and site qualifications. Some of it has to do with study conduct. Do they have their SOPs? Is the sponsor personnel CVs in the master -- along with is the data recorded and all the other things that we've talked about.

One other means of surveillance is a database lock. So again, it's one last look from the time of last patient out of the study, there's generally around a six-week period, sometimes a little less, sometimes a little more, where the monitors go back again, sponsor data management people go back in again, look one more time at the data, look one more time to make sure that everything is entered where it should be, that there's no discrepancies before we lock the

Page 319

database.

It's only when all that has been resolved, when all queries to the sites have been resolved, all AEs have been reported, that we will lock the database, and then we actually can begin the analysis. It's a frustrating time for a sponsor and for me because you know it's done. You know the data is out there, and I can't see it.

I'm going through that right now where we finished a study actually six weeks ago, and we just locked the database yesterday so it's okay. So I will be looking at my email to see if I can see results.

Well, there's medical surveillance, and again, medical monitoring is a key bit of surveillance that goes on during a study. Number one and foremost is maintaining safety of the participants in the study, but then also trying to ensure that high-quality safety and efficacy data is collected. And it's not just from the medical monitor or the sponsor; it's the entire team. It's the monitors that go out to the site. It's the

Page 320

operations folks, the statisticians and so on.

So I think everyone knows the medical monitor, generally these days outsourced to the CRO, is the boots on the ground. They're the ones interacting directly with the sites, receiving the queries about inclusion/exclusion criteria at 3:00 in the morning, having the direct contact, talking to them on the phone, discussing an abnormal laboratory value and what that means, generally can this patient come into the study or can this patient stay in the study, those types of things.

But again, there's a monitor for the monitor, and that's the study physician, generally internal, working for the sponsor, not always, and addresses the issues that the monitor needs to escalate where there wasn't general agreement between the medical monitor and the site, oversight of the medical monitor. And then some key features actually beginning to look at some of the trends, not waiting for the end of the study but beginning to look at some of the blinded data, particularly the safety.

Page 321

Just like there's a vendor management plan, there's a medical monitoring plan created jointly by the clinical study physician from the sponsor and the medical monitor from the CRO. They're partners in this study and looks at some of the issues that we've talked about so that everyone knows before the study starts who's going to be responsible for what.

Then, as I've mentioned, there's safety medical monitoring. We actually have the ability -- and Nat showed one program. Teva has a program. I think a lot of sponsors have different monitoring programs -- where at a given time in an individual study, we can track blinded safety data looking at what the lab results are.

Is there an alarming trend for elevated liver function tests? Is there something that we're particularly looking for? Do we know that there's a risk of elevated liver transaminases and we want to track that particularly? Or adverse events of interest, we know there might be a risk of dermal reactions, so we want to see. And then

Page 322

looking at population trend, looking across the entire study mainly for safety of patients in the study, and we don't want to continue a study where there's an alarming increase in an adverse event.

The data on the right is blurred for a reason. You're not supposed to read it. It's just an example of some of the things we see.

Periodic reports, the study physician is looking at some of this blinded data and answering some of those questions. Are there new adverse events we didn't predict? Are there things we didn't expect that we should be aware of? What about the AEs of interest? What about the study conduct? Are protocol violations occurring more at a particular site? Is there some kind of suspicious medical history? And getting back to some of the things we talked about earlier in terms of patient conduct but also site conduct.

I think everyone's familiar with IRBs and ethics committees, but a kind of surveillance that happens before the study starts and then continues during the study. We may go to the IRB or the

Page 323

ethics committee with some alarming blinded data we're concerned about, or we may for some studies have a data monitoring committee, an independent committee that can look at the data and tell us without us being aware, without breaking the blind, you guys have a problem with some of this. Again, all out to protect patient safety.

So then just a few surveillance questions. Should source document verification be a mandatory requirement? I may have hinted at what I think, but again, curious to hear.

How do we ensure the proficiency of the monitors? Who's monitoring the site monitors? How do we know that they're doing what they should? I don't. We send them out there, but I'm going to assume that some are better than others. How do we know?

We know that DMCs are out there to evaluate safety. Do we have any kind of independent board or group of people that can monitor study conduct? I mentioned to you that we may do some of that on our own looking at protocol violations, but do we

Page 324

need an independent group to do that? I don't know.

So I think with that, I will stop and used up enough time so there's no time for questions today. Thank you.

(Applause.)

MODERATOR: We have time for about five questions. No.

(Laughter.)

MODERATOR: So I'm delighted to introduce our next speaker who has come a long way to talk to us about central statistical monitoring. One of the papers you have in your packet was written by her and her colleagues. So Amy Kirkwood is a biostatistician and senior research associate at Cancer Research U.K. and University College London Cancer Trial Centre.

So as I said, she's going to speak to us about some of the details of the things they do for central statistical monitoring.

Presentation – Amy Kirkwood

MS. KIRKWOOD: Thank you.

Page 325

So I'm Amy Kirkwood. I'm a statistician at CRUK and UCL Cancer Trials Centre. So central statistical monitoring is something I started to look at a few years ago when there didn't seem to be that much interest in it. It seemed to be a topic that had been discussed for about a decade or longer, but no one had really done anything with it. So we thought, is this something we could do for our trials?

I'm going to discuss today what we've developed within our trial center, how we applied it to our trials and what we found, what we're going to be doing with it in the future, then some more information on what other people are doing with these sorts of ideas.

It's become a much more popular topic in probably the last sort of three to four years. It seems a lot more people publishing papers on it and applying it to real trial data. But there's still things that we haven't answered, and I'll go through some details on questions we still need to answer about central statistical monitoring.

Page 326

So a little bit of detail before I go on about our trial center. So we run academic clinical trials. None of them are licensing studies. So some of our trial data may be used for licensing applications, but none of them were directly licensing studies. So that means none of our trials have the kind of monitoring that was described in the previous talk.

Patients don't receive any financial compensation for entering any of our trials, and the sites don't benefit directly, either.

This is something that we discussed earlier, but again, all of our trials are run through the NHS. So there's also no benefit for the patients for going into the trials in order to get free medical treatment. They would be getting the standard of care whether they entered the trial or not.

Until last year, all of our trials collected data using paper CRFs, and we've only just started to move into ECRF. So that might change the way we monitor data in the future.

Page 327

Our databases have minimal in-built validation checks. This is partly due to the fact that we're an academic unit and it takes a lot of resources to program these in, but also because it slows down the programs we use.

We use a risk-based approach to monitoring as recommended by the MHRA, which is the body that governs INP trials in the U.K. So this means our early phase, high risk studies. So things that use unlicensed drugs or advanced therapies, will get quite a lot of monitoring, whereas our later phase trials, particularly ones that use licensed drugs, may have very little or no onsite monitoring.

When we do these onsite monitoring visits, we look at various things, so drug accountability, lab monitoring consent, and some cases, we will do source data verification.

So the aim of central statistical monitoring is really to cut down on source data verification. So this is kind of going to replace the last talk about the views on source data verification.

So we thought of the reasons to look at

Page 328

source data. So they were data errors, which we're talking about just unintentional transcription errors, typos, et cetera. Procedural errors, so these are things we classified as unintentional errors made by the site where they really just hadn't understood the trial protocol. And finally, fraud, so this would be intentionally making up patients or making up patients' data.

SDV is a very common activity, and there was a paper in 2011, which may answer some of Richard's questions from the previous talk. They surveyed lots of different types of organizations that performed trials, and they asked them about onsite monitoring.

So 77 percent of their respondents always performed onsite monitoring, and at onsite monitoring visits, SDV was always performed by 74 percent. And not so surprising, I guess, it was more common in pharmaceutical organizations than it was in academic institutions like ours.

There are a few studies that then show that SDV might actually not be that useful The Bakobaki

Page 329

paper looked at errors they found during monitoring visits, and they decided that 28 percent could have been found during the data analysis, so this would be the standard sorts of things that you'd look for when you're analyzing the data. And they would have picked up those errors at that point. And 67 percent could have been found through other centralized processes, so without actually going to visit the site.

Another paper by Sheetz in 2014 said that in over 1,000 trials they looked at, only 3.7 percent of ECRF data was actually corrected, and only 1.1 percent was through SDV.

This was something else that was shown by Tudor-Smith in 2012, which found that the majority of the SDV findings were random transcription errors and had very little impact on the main conclusions, and actually missed four ineligible patients. And this is something else that Grimes talked about in their 2005 GCP guidelines, that if your source data matches the data that's being sent in, it won't get picked as discrepant even though

Page 330

it might be incorrect in some other way.

So we thought what if we could do statistical methods centrally to try and reduce some of this source data verification? So the idea, it would save time on site visits or even cut down the number of site visits. And if you still go to site, you could spend that time doing things like staff training and other things that you can't do remotely.

When I started to look at this a few years ago, there'd obviously been an interest in suggestions that this is something that we should be doing, but people hadn't really applied it to the sort of trials we were running. And no one really developed any software to do it.

This is a selection of the references that were around at the time I started this project. The first three sort of covered lots of different possible methods to look for fraud and data errors. A lot of these we used when we developed our programs.

The second, although they applied it to real

Page 331

data, it wasn't necessarily particularly applicable to our trials. Some of the were questionnaire data, some were in animal models. And they didn't necessarily use sort of all of the techniques in the previous papers. They just picked and chose one or two.

The aim of our project was to develop a set of programs that could be run in R that would perform most of the sort of common checks that had been suggested. So these are things that wouldn't easily be done by the clinical trial database while the data was being entered, and we wanted to create output that was hopefully straightforward enough to be understood by a non-statistician.

So all the checks we selected, we split into two categories: things that were at the trial subject level or at the site level. The checks at the subject level are looking for possible data errors within individual patient's data. A lot of these things are things you would do when you do your data analysis at the end of the trial anyway, so checking the order the dates fell in, were

Page 332

patients randomized before they were treated, were they treated after they died, sort of obvious things, but we wanted to set up an automated system for this.

We also looked at whether the dates fell on weekends or national holidays. And obviously things like dates of death or dates of adverse events could fall at any point, but in our center, there's no way a patient would be randomized on a weekend or a national holiday because we're closed.

The same for a lot of radiotherapy treatment and chemotherapy in the U.K., a lot of it won't take place on weekends and national holidays. So this could be an indication that the date is wrong.

The next little test we looked at were methods for detecting outliers. Again, standard things that you might do while you're analyzing the data, but we wanted to find some way that we could do all of these things quickly and easily as the trial was going on. So hopefully data errors could be corrected as we go along rather than waiting till the end of the trial.

Page 333

All of these checks aimed to find recording and data entry errors, but the data checks may also detect fraud if people are being sloppy with how they made up their data. And procedural errors might be picked up either because we'll find things like patients being treated before they were randomized, or if it's a lot of outliers, that might indicate inclusion or exclusion criteria that have not been followed.

We could also sort these in our output by site. So if we've got sites that are continually sending in data with errors, then that might be a reason to go and talk to them or give them further training.

When we come to checks at the center level, here, we're aiming to flag centers which are discrepant from the rest by looking for unusual data patterns. So these are mostly aimed at detecting fraud and other procedural errors. I'm going to go through examples of these in the next few slides.

So one of the things we looked at was digit

Page 334

preference, the idea being that humans really are poor random number generators, and that if you're making up data it may not follow the same sort of distribution as real data would.

What we did is we compared the leading digit, the distribution of the leading digit in each site with all of the other sites put together. And this would flag any site where there seemed to be a difference in the number of ones, the number of twos, number of threes compared with each site and the rest of the sites put together.

We also looked at rounding. So we could either look at the tailing digit and use a similar method to the method I just described, and there, you'd be looking for an increased number of zeros or 0.5s, or we'd probably run a graphical method, which would show where integer values had been recorded.

So in this case, if you've got a non-integer value, the curve takes a step up, and if you've got an integer value, it forms a horizontal line. So something like site 63 with the two little

Page 335

horizontal steps, you wouldn't worry about it. But if this was an important variable which had to be recorded to a certain level of accuracy, if you got something like site 68 where they're continually recording integer values, it might be a reason to talk to them.

Another thing we looked at was inliers. So inliers are basically exactly what they sound like. They're the opposite of outliers. So an outlier is a point that lies far away from the rest of the data, whereas an inlier is a point that lies close to the center of the data.

So here, you take a selection of continuous variables a one in CRF, and you calculate the distance between the mean and the point on each patient and you sum that across all the variables. If you plot this on a log scale, what you're looking for are points that fall at the bottom of these plots. As you can see in the first plot, we've got one that's circled in red.

We found when we faked data ourselves to be close to the mean that could be detected, but we

Page 336

had to be -- we felt you had to look at both the plots and the points, which had been selected by the program, as if you added more than one of these patients, they often wouldn't be picked up.

So the idea here is you're falsifying data, you want your data to look as believable as possible, and you want it to look like the other patients. So you're going to make it as similar to the other patients as possible, and this might be a method to do it.

The next thing we looked at was checks of the correlations structure. Even if people were good at falsifying individual data points, they may not get the interactions between those variables correct. Here in the graphs, you can see all of the continuous variables on one CRF, and it plots the correlation between the pairs of variables.

So if there's a perfect positive correlation of 1, it's represented by a black square, which is why you get a black diagonal down the middle. If it's a perfect negative correlation of minus 1, it would be represented by a white square. And

Page 337

everything else is like a shade of gray in between.
So what we're looking for here are sites that don't appear to have the same correlation structure as the rest of the sites.
The two examples in the middle are sites that we created. The first one on the left-hand side, we created by generating patients that had values to each variable that lay close to the means. So what this showed is you ended up with a kind of overall light gray structure. Although it doesn't look that different to the site above and below, you'll notice there's some strong correlations that are missing.
In the second example, we generated a fake site by just putting random values from all of the patients, and what you get there is a site that looks strikingly different, that you've got strong correlations that don't exist in the other sites.
These are tested to p-values using simulations which compared the correlation structure to the overall correlation structure for all of the other sites put together.

Page 338

As with the inliers test, we felt there's another thing we needed to look at both plots and p-values. Something that looks quite strikingly different, like the site on the right, wouldn't have nearly as small a p-value as the site on the left just due to small numbers of patients.
Another suggested method was to look at the variance within repeated measurements data. So there's a suggestion that if people falsify data, they don't get enough variance when they're looking at repeated measurements data.
One of the earlier references I had was from an animal trial where they had detected fraud by looking at the variance in the measurements collected. All this does is plot all of the patients' values for a particular continuous measurement in a line across, and it highlights patients with very large variances in shades of red or very small variances in shades of blue. So a very large variance might indicate that you've got data errors or outliers, whereas very small variance might indicate fraud.

Page 339

What we can see in this top plot in site 25, a three-patient chain in blue. So this was actually data that we had given to an independent statistician and asked him to falsify some patients for us. So these were three falsified patients that he had created, obviously not that well, and we'd obviously not given him any warnings about the kind of tests we were doing.
One of the most important sort of monitoring activities in clinical trials is to look at rates of adverse events. So we wanted to find some sort of way to monitor this. This is really to detect any sites that might be underreporting adverse events.
Here, we created a severe adverse event rate for each site as the number of patients who had an SAE reported, divided by the total number of patients in that site and a measure of the time in the trial. So we plotted this rate against the total number of patients, and what we'd be interested really are the centers which fall in the bottom right-hand corner. They have quite a large

Page 340

number of patients but still a low rate. The lowest 10 percent of rates are picked out automatically and shown as black squares there.
The other thing is this output is also the number of SAEs the site has in total. So this might be an indication that, okay, they've only had X percent of patients having an SAE, but they have reported a lot of SAEs for those patients, so they may be on top of their reporting. We also felt you need to look at how this rate compared to the overall rate.
We also thought this could be adapted to look at incidence reports where you'd be interested in sites with high rates of incidence and low numbers of patients.
This is something that we found pretty useless, but it creates quite a nice picture. This was an idea that a lot of the other authors suggested was some sort of graphical representation of the means of different continuous variables, so you could see if some sites stood out compared to the others. And this is a technique called

Page 341

Chernoff face plots, the idea being that humans recognize differences in faces better than they do other graphical techniques.

Every feature on the face is controlled by a different variable, so the height of the head, the width of the head, the height of the eyes, the width of the eyes, et cetera. We found it to be, though, as I said, kind of useless.

What variable controlled which feature had a big impact. So, for example, site 11 here that stands out quite a lot, that was the height of the face being controlled by a variable that had one massive outlier in that site. And as you can see, if we deleted the outlier, it went back to looking much more like the other sites.

You can see -- or you might not be able to see because it's quite small, but site 48 has really huge eyes, but it doesn't stand out because it's indicating there was quite a big difference in whatever variable created that. But you just don't really notice it. So this is something that we're still thinking about the best way to approach.

Page 342

Another suggestion was star plots, which again, we found pretty much uninterpretable.

So what did we find in our trials? So we looked at three trials, a phase 3 lung cancer trial where we'd already cleaned the data and had it published, and another phase 3 lung cancer trial, which was in follow-up but the data hadn't been completely cleaned or analyzed.

So in the first trial, we found some data errors which hadn't been detected. So we couldn't go back to site, but we could go back to our paper CRFs and we could see that some things were clearly wrong.

Some outliers, which were possible errors, though none were used in the main analysis. Some patients who were treated before randomization, though, after discussions with the trial staff, it turned out that this had been known. And because of the nature of the trial, patients who had only just started treatment, as it was standard care plus another drug, were allowed into the trial in this case.

Page 343

One site had a very low rate of SAEs, which we probably would have wanted to look into, but obviously at this stage, we couldn't. We did have some failures in some of the center level checks, so these are the checks for fraud.

For example, the digit preference was flagged by one center, but we looked at this center in another CRF, and we didn't see the same pattern. So we didn't have any concerns that there was actually any fraud in this trial.

We found very similar results for trial 2, though, unsurprisingly, we detected a lot more data errors and a lot more potential outliers.

The third trial we tested this on was a phase 3 trial of biliary tract cancer. This was another trial that had been cleaned and analyzed, but we used this trial to generate our own fake data to see what could and couldn't be detected.

So unsurprisingly, when we generated data that fit the assumption of the programs, we were able to detect it in most cases, though the amount of fake data obviously determined how easy it was

Page 344

to detect. The independent statistician who added data to this trial, some of his data was also at the sites, but his data were also picked up by several programs.

So how is this going to be put into practice in our CTU? So we will choose a test to apply based on the size of the trial and how far along the trial is. So this is one problem with these sorts of methods is that you can only -- particularly the methods for detecting fraud, you can only apply when you have a certain number of patients in your site.

So if you've got small early phase trials, they're not going to be very helpful, or if you've got trials that haven't yet recruited that many patients at each site, they may also not be that helpful.

So data's going to be checked at appropriate regular intervals. We've done some SAE monitoring using these programs so far, and we went for every six months. But I think it would probably depend on the rate of recruitment into your trial. These

Page 345

would be set up by the trial statistician, and then they should be able to be run pretty much automatically.

Data errors will be discussed with the trial manager or the trial coordinator to see what we should be taking back to site and which we shouldn't. If there's anything more worrying, so this is anything that suggests procedural errors or fraud, these will be discussed with the appropriate trial staff and also our regulatory department.

So the SAE monitoring I mentioned, this had been applied to another lung cancer trial in our center. And after two checks on the SAE rate -- there was one site on the third check which had had pretty low rate all the way along. And we thought, okay, so we'll go with the gentle approach and send them just a nice email that sort of said, you do know that you're meant to be reporting all of the SAEs. There's none that you're not reporting, are you? And they came back, oh, no, no, of course not.

By the fourth time we looked at this, they

Page 346

had gone from eight -- they'd got one extra patient and gone from eight, I think, patients with SAE to 21 patients with an SAE. So we don't know whether that was our little nudge or not, but it will be interesting to monitor the changes you get when you do contact sites and see if it does alter their behavior.

So as I said, this is something that's become a lot more popular in the last few years. So since we finished working on our paper, these are a selection of references that have been published.

It seems to sort of split people into two groups. There are papers that tend to use sort of all of the data or systems that use all of the data, and then those that take specific key variables and monitor them, and I'm going to go through two examples.

One company that has spent a lot of time and research in the last few years is this company called Cluepoints. So this is a company started by Marc Buyse who wrote one of the original papers on

Page 347

how to detect fraud and data errors in clinical trials. So they offer their services running CSM to CROs, academic staff, and pharma companies.

So they apply similar methods to the ones I've described, and they also look at things like rates of missing data. All of their tests produce a p-value. I went to a talk taught by one of the authors of this paper a couple of weeks ago, and they said for a phase 3 trial, they could have 100,000 to a million p-values generated by their programs.

They get this huge matrix of p-values, and they use a principal component analysis on it to try and pick out centers that stand out. So that's what these two plots show. So this circled site, site X, was a site where they knew there had been fraud. And what they were looking for is sites that fell far away from the origin because that suggested they're different from the rest. And as you can see, site X does fall far from the origin.

But they also have circled centers D6 and F6 in one of the plots and D1 and E6 on the second

Page 348

one. And they do mention in their paper, if this is an ongoing study, you'd want to go and talk to these sites as well. But there didn't seem to be any data about whether there was any irregularities at these sites or not.

The second type of use for CSM is more sort of targeted with just specific key variables. So Oxford has done some work with things they call key risk indicators. This paper developed models using data where they knew there had been fraudulent data. So this was a trial for the POISE trial. They knew there'd been nine sites with falsified data.

So they used again very similar methods to the ones I've described to build risk scores. So these were three variables in each model that would be able to predict whether the site had falsified data or not.

They found these risk scores could discriminate between their fraudulent and their validated centers very well. So they had a very good area under the curve for those. And then they

Page 349

tested these scores on another trial in the same disease area but very similar design of trial where they had had a lot of onsite monitoring. They were sure there had been no fraud.

So the false positive rates in this second trial were very low, so that means it was very few sites being picked as possibly having fraud where they shouldn't be. But they didn't have another trial with fraudulent data to apply these to, so they don't know how well it would be able to pick up fraudulent data in another trial.

So it seemed like an interesting method, but only one might work in this disease area where these particular data points were reported.

So what are the advantages over SDV for central statistical monitoring? So all data could be checked regularly, quickly, and we think cheaply. We'd hope the data errors would be detected early, which would reduce the number of queries needed at the end of the trial.

Procedural errors are more likely to be detected at the end of the trial when they can

Page 350

still be corrected. So I suppose those last two, there are more advantages over not doing this rather than advantages over SDV. And every patient would have some form of data monitoring, whereas at the moment, at least in our center, only a small proportion have onsite monitoring. We'd hope this might be able to pick up anomalies, which existed in the source data as well

So the disadvantages, as I mentioned, some of these methods aren't reliable when there's only a few patients in each site, which isn't surprising, but this could also be an issue particularly early on.

The programs find data errors can be used on all the sites and again, early on, but programs to detect fraud will only be able to be used once you've got reasonable number of patients in each site. And some of the methods are definitely somewhat subjective.

So what other research is needed in this? One thing is how much does it cost. We think it might be cheaper because it might save money on

Page 351

site visits, but no one's actually looked into the cost of implementing these and interpreting the results. This is the time of the statistician or someone else who's running these. And in a center like ours which doesn't do a lot of onsite monitoring, might you actually end up with more for cause monitoring visits which will cost more money. But the most important thing is how can be it validated. How can we assure that sites which aren't flagged didn't have any falsified data?

One study which is trying to do this is a trial which is being run in the U.K. called the TEMPER trial. What this is doing is it's going for the few key variables idea. It's selected specific triggers to look for in a site, and if any site triggers these, it will have a monitoring visit. It then matches this site based on the number of patients it's recruited and the time it's been open to another similar site in the trial, which wasn't flagged, and it will also go and visit that.

What it's aiming to do is to show a 30 percent difference in the numbers of critical or

Page 352

major findings between the trials that were flagged and the trials that weren't. This would be a possible way to test some of these methods. And I don't think anyone's planned a similar study using the kind of methods that we're using or the methods suggested by Cluepoints, which look at a lot more of the data, but that might be something useful to do.

Another possibility is to run a trial which looks at central system monitoring and full onsite monitoring to see what is picked up by each. The problem with doing that is that if sites know they have full onsite monitoring, they may behave differently than if they didn't. And you may not be able to -- if things are changed as you go along during the trial, because of findings in either method, you won't know whether they would have been picked up by the other method or not. So this is something that definitely needs more work, and it needs more consideration of how we can test this.

Finally, there are further details in the paper that I wrote on this subject, which was

handed out in the background reading.
(Applause.)
MODERATOR: I've just been informed by the person who was number 11, I think, on the Chernoff face plot, Bob --
(Laughter.)
MODERATOR: -- that we're going to push Paul's talk to 8:00 tomorrow morning. Did you tell Paul that by the way?
MALE SPEAKER: Yes.
MODERATOR: Okay.
MALE SPEAKER: Paul was very actively involved in the decision.
MODERATOR: Okay. So I guess we'll end now. We have dinner here at 7:00.
FEMALE SPEAKER: We do not (inaudible) dinner in --
MODERATOR: We do not.
(Housekeeping.)

Adjournment

MODERATOR: Okay. So thank you.
(Whereupon, the meeting was adjourned.)

B

C

D

E

G

H

I

N

O

P

Q

R

S

T

W