ACTTION

Measures of Outcome for Stimulant Trials (MOST)

March 25, 2015

A Matter of Record

(301) 890-4188

ACTTION

MEASURES OF OUTCOME FOR STIMULANT TRIALS (MOST)

Wednesday, March 25, 2015

Hilton Rockville Hotel and Executive Meeting Center
Rockville, Maryland

C O N T E N T S

C O N T E N T S (continued)

C O N T E N T S (continued)

Page 5

P R O C E E D I N G S

Welcoming and Introductions

DR. STRAIN: Let me introduce myself. I'm Eric Strain, and I have the dubious distinction of chairing this meeting, and I want to welcome you all to it. This is the Measures of Outcome for Stimulant Trials meeting. And I'm going to explain a little bit of background to it over the next couple of minutes, but won't take long.

Before I get started, let me say that -- actually, Bob Dworkin and I are co-chairing this meeting, but Bob -- I think I can tell you this. Bob has a kidney stone that he developed in San Francisco and is not allowed to travel at this point. I think he's getting a procedure done or he's going to be evaluated for a potential procedure.

So it's just really striking the extent that some people will go to, to avoid coming to a meeting I think. That's why Bob is not going to be here today or tomorrow. And he sends his regrets. He really did wish he could be here.

Page 6

A few housekeeping things, just standard things. Please sign on [sic] to your cell phones. The microphones here are voice activated, so please speak into them if you do have a comment to make. And I hope that people will engage in discussion over the course of the next couple of days.

This meeting is being recorded. There will be a transcript generated that's put on the ACTTION website, so that you're aware of that. That's been pretty routine for these meetings. And we will have breaks that actually will be in this room right over here to my left, to your right. So they're going to be setting up coffee and things in that room a little bit later this morning, so we won't need to go back down the hall to the other room to get replenished. Lunch will be held in the Roosevelt Room located in the lobby level. We'll explain that to you later

Check out is at noon tomorrow, and taxis can be ordered for your return to the airport if you're flying. And you can check in with Valorie. Valorie, you want to wave your hand? If you

Page 7

haven't met Valorie in the back, if you have any questions about those sorts of things -- and you should have signed in when you came in this morning. There was a sign-in sheet at the front desk, right outside the door here.

Let me just give you a few orienting things about this meeting. You should have received this, but just to remind you, the goals of this meeting are to review work that's been conducted validating outcome measures for clinical trials of stimulant-use disorders and to review related target substances that have successfully developed clinically meaningful outcome measures for substance-use disorders. And finally, to identify research agenda for the development of a tool and related outcome measures that would be clinically meaningful for stimulant-use disorder clinical trials.

Those were the things that were defined to us. And the structure of the meeting is that we decided at the beginning we wanted to keep this relatively small, somewhere between 30, 45 people

Page 8

max, so that's where we're at. I think we'll have about 35 or 38 people. Some people will be arriving later this morning. There are a couple of people who will be here for tomorrow only.

The presentations, as you probably see with the agenda, tend to be -- there are presentations with a moderator, and then there's a discussant afterward who will hopefully bring together and draw out some of the salient points from the presentations. And as I mentioned before, we are recording the proceedings of the meeting.

MALE SPEAKER: Will slides be available?

DR. STRAIN: I think they are. I think the slides go up on the website afterwards. Yes.

Some people have asked me -- and I have asked -- how does this all work? This is sponsored by ACTTION, which actually comes out as AAA-CTTION, because there's a bunch of A's and T's in ACTTION, which is the Analgesic, Anesthetic, and Addiction Clinical Trial Translations, Innovations, Opportunities and Networks program. I'm not sure if Bob Dworkin or Dennis Turk -- one of them loved

Page 9

acronyms -- hence, MOST -- and ACTTION is a public-private cooperation that works to optimize clinical trials methodologies, and has included academicians, people from the FDA, people from NIH, as well as industry.

Its historically worked in the area of analgesics and of pain, but they've recently taken interest in the area of addictions. They had a meeting looking at the use of terminology, especially misuse terminology, and produced a paper out of that. The primary mode of ACTTION -- and I couldn't resist that pun -- is through expert panel meetings and reports, such as what we're doing here.

This is really run out of the University of Rochester primarily and comes from funding from a variety of federal as well as private sources that seeks to produce something that will move the field forward, for example, in terms of either definitions of terminology or defining research needs of the field.

The planned outcome for this meeting is a

Page 10

paper and a peer-reviewed literature that reports on the outcomes, so it might be on the current state of the affairs, the needs, the research agenda, and going forward. And Brian Kiluk, who can wave his hands, will be taking the lead on writing up the results of this meeting. And that's very in keeping with other ACTTION efforts.

So let me stop there. Any questions about any of that?

(No response.)

DR. STRAIN: I think it's pretty straightforward I hope.

Again, the meeting is structured with time for discussion, time for reflection. I realize we all may not have the answers, but I hope that we can at least help to formulate questions and ideas as to what needs to be done to help us get those answers.

Let me see if there anything else I didn't touch on. I think the only other thing is a housekeeping thing, bathrooms. Bathrooms are down the hall and to the left. Other than that, I think

Page 11

we're set to go.

So I would like to, next then, introduce Ashley Slagle from the FDA, who's going to be talking about approach to outcome measure development or selection, a regulatory perspective.

Ashley?

Oh. And I should mention, I'll serve as the discussant for this pair of talks since Bob Dworkin's not available.

Presentation - Ashley Slagle

DR. SLAGLE: Good morning. Well, thank you very much. I appreciate the invitation to speak today. I'd like to give a little bit of background and sort of set the stage for the discussion today to share a regulatory perspective on the approach to outcome measure development or selection. Of course, the views expressed in this presentation are my own and don't necessarily represent an official FDA position.

Before we get into the details of measure selection or development, I want to step back just a little bit and think about the broader context of

Page 12

what we're trying to do in clinical trials. So ultimately, we seek to evaluate treatment benefit; that is, that a drug has some positive impact on something that is important to patients: so how long they live, how they feel, or function in daily life. We then have to weigh those benefits demonstrated and quantified in clinical trials with known risks of the products in order to make drug approval and labeling decisions.

We use outcome assessments to determine whether or not a drug has been shown to provide benefits to patients. One of the most important aspects of drug development is how treatment benefit, or sometimes referred to as clinical benefit, is measured.

There are different types of outcome assessments can be used to evaluate treatment benefit. Of course, there's survival, but in many contexts, survival may not be an appropriate endpoint in a clinical trial, either because the condition does not impact survival or because the trials would need to be prohibitively large or long

Page 13

in order to detect survival benefits. So we often turn to other types of outcome assessments.

Then there are clinical outcome assessments of which there are four types: performance outcome measures, and then outcome measures that are reported by either clinicians, other observers, or patients themselves. So it's important to note that what separates clinical outcome assessments from the other types of assessments are that they rely on human judgment, motivation, or participation.

So it might be that we're interested in symptoms in a population that can report for themselves. In this case, a patient reported outcome assessment would be most appropriate. If clinical judgment is needed to interpret an observation, then a clinician reported outcome assessment is most appropriate.

If an observable behavior in daily life is being assessed in a population that can't report for themselves, then an observer reported outcome assessment would be appropriate. In some cases, we

Page 14

want to observe an actual demonstration of some activity in the clinic, and in that case, a performance outcome measure could be used.

So of course patient reported outcomes have gained increasing attention over recent years, and this is really a good thing because including patient reported outcomes in clinical trials is really useful to help us understand directly how products impact how patients feel and function in daily life. But keep in mind that there are other types of clinical outcome assessments that are also very important and can be considered for use in trials. And then there are surrogates, which are usually biomarkers that are intended to be used as a substitute for how a patient feels, functions, or survives.

Let's talk a little bit more about treatment benefit. We think about treatment benefit in terms of direct evidence and indirect evidence. Direct evidence of treatment benefit is derived from studies with endpoints that measure survival or how patients feel or function in daily life. Indirect

Page 15

evidence of treatment benefit is derived from studies with endpoints that measure other things that are related to how patients survive, feel, or function.

It might be helpful to think in terms of a continuum of direct and indirect evidence of treatment benefit. Direct evidence of treatment benefit is gained from actually measuring what is meaningful to patients in their daily lives.

For indirect evidence of treatment benefit that does not actually measure the clinical benefit directly, we need to have some evidence of the link between the indirect assessment and the meaningful benefit. So depending on how indirect something is, the more evidence we might need to understand that link.

For example, we consider performance measures like the 6-minute walk test to be somewhat indirect because they are not measuring how people feel or function in their daily life, but are intended to closely approximate how patients feel or function in daily life. With these types of

Page 16

performance measures, it's important to understand what the performance test is actually measuring and what it is intended to represent. So what does a score change on this indirect measure mean in terms of meaningful treatment benefit in patient's daily life.

Biomarkers are at the far indirect side of the continuum, so therefore, we need very strong evidence showing that the biomarkers predict some clinical benefit; again, how patients feel, function, or survive.

Surrogates within existing and well-established link or prediction of clinical benefit can support endpoints that are used in traditional approval. So an example would be something like blood pressure. Biomarkers or surrogates without that existing evidence that they are linked to meaningful clinical benefit, but are reasonably likely to predict clinical benefit, might be able to support approval through the accelerated approval pathway with the requirement that post-approval studies are completed to then

Page 17

confirm that link between the biomarker and the expected clinical benefit.

We want to include the patient perspective and evaluate patient reported efficacy outcomes in clinical trials. How do we determine what's best to measure? So we need to consider within our selected patient population what concepts are relevant to how patients feel and function.

For drug approval and labeling, we also need to think about how closely related those things are that we might measure to the disease and the treatment. This slide helps to define those things that might be more proximal to the condition and treatment, like poor symptoms and aspects of functioning, and those things that are more distal like health related quality of life.

This doesn't mean that the more distal concepts are less important. It just means that there are many more variables that might impact those concepts in addition to the disease and the treatment. So the farther we move to the right on this diagram, the harder it becomes to detect a

Page 18

treatment difference or to interpret any treatment difference that is identified when you measure these things.

Of course, the distal concepts are considered for measurement in clinical trials to support labeling claims. We need to ensure that the variables that contribute to these concepts are also measured so that we can interpret trial results. For example, if we wish to measure health-related quality of life, we need to make sure to assess symptoms, adverse events and toxicities, and all of the impacts that can contribute to health-related quality of life, including general psychological functioning, physical functioning, social functioning, and so on.

Measuring these more distal concepts becomes increasingly difficult, and especially in patients with multiple co-morbidities that might also impact things like health status and quality of life that might be unaffected by our treatment under study. So that's why we often recommend that more specific

Page 19

aspects of quality of life that are closely related to the condition and treatment, like symptoms or proximal daily functioning impacts, be the target of endpoints and clinical trials, rather than the broader concepts that are of course important but are harder to measure, interpret, and to show a treatment effect on.

From the regulatory perspective, it's necessary that drug developers document substantial evidence of treatment benefit from adequate and well-controlled clinical trials. The regulations also specifically indicate that the methods of assessment of a subject's response should be well-defined and reliable. So this is important. It means that well defined and reliable become the key criteria by which the FDA evaluates outcome assessments to document evidence of treatment benefit.

So when is a clinical outcome assessment well defined and reliable and appropriate for use in adequate and well-controlled studies? Well, when we're measuring the right thing in the right

Page 20

way, in that population, and that the score that quantifies that thing that we're measuring does so accurately and reliably so that the effects that are seen in the outcome assessment can be interpreted as clear treatment benefit.

We refer to the FDA patient reported outcome, or PRO guidance, that describes good measurement principles that might be considered to evaluate whether a measurement is well defined and reliable. This guidance was developed specifically for patient reported outcomes, but many of the principles are appropriate and apply to any clinical outcome assessment type.

This guidance provides an optimal approach to patient reported outcome development, but we understand that flexibility and judgment are needed in order to both meet the regulatory standards as well as the practical demands of drug development.

Specifically, when we evaluate whether an assessment is well defined and reliable, we evaluate the tool's measurement properties. First and foremost, we consider content validity. After

Page 21

content validity is established, then we consider other measurement properties, including construct validity, reliability, ability to detect change, and then guidelines for interpreting meaningful change.

So what do I mean by content validity? We often use this term "content validity," and I think eyes glaze over sometimes. So I want to just talk a little bit about specifically what we mean. It means what are we measuring? Is that the right thing to measure in that population? Does the patient understand the items and respond in the way intended? And then when we combine all of the items from a questionnaire into one score, what does that score represent?

As regulators, we put a big emphasis on content validity because we need to ensure that when we see score change on an assessment, we can determine what that score change means. And importantly, that we can describe that score change in labeling in terms of meaningful treatment benefit in a way that it's not potentially false or

Page 22

misleading.

So we do recommend involving patients, using focus groups or individual interviews, to help develop the assessments that are planned for use in clinical trials. And this can help ensure that we're measuring the right things and to help us figure out the best way to ask the questions in the trials.

Another key consideration is how to interpret what is meaningful change on an outcome assessment. So often statistically significant changes alone are not fully interpretable. So if we see a very small change in a score that is statistically significant, we have to think about whether that amount of improvement is meaningful to the patient population, and weigh the amount of improvement or benefit against risk.

So we really need to think very carefully when weighing risks and benefits of drugs. And of course, these aren't unusual decisions. It's really the job of the FDA to incorporate our regulations, science, and judgment to weigh

Page 23

quantified risks and benefits to make approval decisions. But this is also where patient input can be very useful to help us understand what is a meaningful change, amount of change on an assessment, and how do patients weigh those risks and benefits?

It's possible that for some things, a cohort of patients with one condition might think very differently about what are meaningful changes than a cohort of patients with another condition. So we need to understand and incorporate these considerations.

We've developed a couple of diagrams to help facilitate assessment development or selection. The first is a Roadmap to Patient-Focused Outcome Measurement in Clinical Trials, and this is really intended to illustrate how someone might embark on a sound and orderly instrument selection or development pathway, beginning with the clinical context in which the instrument is intended to be used.

This is not intended to be a hurdle or a

Page 24

checklist. I know there's a lot of information on here. It really is intended to be helpful. It's meant to be a tool that organizes a lot of the things that drug developers are already thinking about, and some things that sometimes they forget to think about. But it really can help inform drug development programs and outcome assessments. I'm not going to go through it in detail. I just wanted to alert you to the existence of this tool on our website, and then drive home a couple of key points from the diagram.

First, it's really important that adequate attention is given to the first two columns. So understanding the condition and conceptualizing treatment benefit before beginning to think about selecting or developing an outcome measure. So this is a common pitfall that we do see. Trial designers and instrument developers haven't paid adequate attention to these things before selecting an outcome assessment, and this can be really problematic for trials.

So we encourage drug developers to discuss

Page 25

outcome assessments early in the drug development program so we can provide advice early enough to help avoid some of the pitfalls that we've seen in the past and improve the likelihood of a successful use of outcome assessments.

There are two elements from the roadmap that I want to highlight that are important when conceptualizing treatment benefit. It's important that we think about what is important to measure and be sure that we measure that appropriate for a given context or context of use of the assessment.

So we've tried to put together a list of common elements that are part of the context of use that might impact decisions about the assessment of treatment benefit. And this list is not perfectly comprehensive, nor will every element apply to every drug development program, but it's useful to give some thought to these elements. Consider the disease definitions: the patient subpopulations, clinical trial design and objectives, and the clinical practice in study settings.

So within the study design objective

Page 26

headings, we include the bullets endpoint definition and endpoint positioning. This is also important in the regulatory setting and can impact our choice of outcome assessments to support endpoints, as well as the level of evidence needed to support the selection of an outcome assessment. Note that outcome assessments and endpoints are not synonymous, but the score of an outcome assessment is used to develop an endpoint definition.

We also think about the following categories in the hierarchy of endpoints. They are primary, secondary, and exploratory. For primary endpoints that are meant to support drug approval decisions, a higher level of evidence is needed to support the selection or development of a particular outcome assessment that forms the basis of the primary endpoint, and then the indication statement and labeling.

Secondary endpoints are generally meant to support the findings from the primary endpoint and can help us better interpret the primary endpoint, or to learn and to be able to communicate more

Page 27

about the drug product and labeling. These assessments will still need appropriate attention, as they could be the basis for labeling claims that may be found, for example, in Section 14, the clinical studies section of labeling. All the assessments need to be well defined and reliable.

Exploratory endpoints might be hypothesis generating, might be used as additional supportive evidence to help interpret the findings from primary and secondary endpoints. But these assessments supporting exploratory endpoints will not be the basis for labeling claims. So these assessments might not need to meet the same level of evidence and documentation to justify their use in clinical trials.

Selecting what to measure and how to place that in an endpoint hierarchy are important considerations. Sometimes multiple measures and endpoints might be used to provide the needed information for approval.

For example, on a recent drug development program for ruxolitinib, a treatment of

Page 28

myelofibrosis, the primary endpoint was a reduction in spleen size. So is shrinking a patient's spleen true clinical benefit? Well, ruxolitinib also demonstrated reduced total symptom score on the Myelofibrosis Symptom Assessment form, which is a patient reported outcome assessment that was used to support a key secondary endpoint.

So in addition to the reduction in radiographic spleen volume, there was also an improvement in patient's symptoms. So this total symptom score was very helpful in correlating the anti-tumor effect with improvements in how patients were feeling and their symptoms. And so ruxolitinib, or Jakafi, was granted traditional approval rather than accelerated approval.

In addition to the roadmap that helps to conceptualize treatment benefit, the second diagram on our website is the clinical outcome assessment wheel and spokes diagram. This diagram identifies the key components of the documentation that would need to be submitted to CDER to support the use of clinical outcome assessment.

Page 29

It really represents the general iterative process of developing a clinical outcome assessment. And I think it's important to note that this type of work has been going on for years in the social sciences, and it's been more recently that we're now bringing this into the world of drug development.

This is a high level view of the wheel and spokes diagram, and a more detailed version is available on our website. For our purposes today, though, I'm not going to walk through this, but I do encourage you to take a look at the website.

The first spoke shows the need to identify the context of use and the concept of interest, so what are you measuring and what's the clinical context?

In spoke 2, you're drafting the instrument and evaluating content validity. And so, this is where patient interviews, or interviews with clinicians, literature, research, can be useful, developing the items, and ensuring that you're measuring what you intend to measure, and that the

Page 30

score really represents what you believe it represents.

In spoke 3, this is the cross-sectional evaluation of other measurement properties. So after you've established content validity, then you can evaluate cross-sectionally construct validity and maybe some test/retest reliability. And then spoke 4 is the longer, bigger, longitudinal evaluation of the measurement properties, and looking at the instrument's ability to detect change and then establishing guidelines for interpreting change.

The final spoke 5 is included if there needs to be modifications to the instrument. So if you're using an existing instrument in, say, a new context of use, and you may need to make some adjustments to the instrument to be appropriate for that new context of use, spoke 5 represents that potential.

I want to share just a few more considerations about outcome assessments, including patient reported outcomes. It's important to

Page 31

remember that available assessments are not all adequate for use as clinical outcome assessments to evaluate efficacy in trials.

So there's no such thing as an instrument that is "validated" for all purposes. Some measures may be used for diagnostic purposes, prognostic purposes, used to select patients for participation in clinical trials, used for epidemiologic or population studies to better understand characteristics or the natural history of the condition, or used to assist in clinical practice decision-making. But assessments used for these other purposes are often not appropriate for use as outcome assessments in clinical trials, at least not without some modifications.

So we've seen challenges across various development programs and understand the nuances of outcome measurement within a regulatory context, so we'd like, to the extent that we can, to share our learnings with drug developers to help ensure the highest likelihood of being able to detect meaningful treatment benefit within trials to bring

Page 32

good drugs to market, and then to provide patients and other stakeholders with important information about those drugs and how patients feel and function in daily life.

We have two pathways to provide advice to those who are interested in outcome assessment development for clinical trials, first within the context of individual drug development programs. So we encourage sponsors to begin these discussions within their individual drug development program as early as possible, even at the pre-IND stage if possible; again, so that if any work needs to be done on the proposed instruments, there's time within what we know are very tight development timelines.

The second pathway is outside of any individual drug development program, and this is through our drug development tool, or DDT, qualification process. And this program, we can work with outcome assessment developers to develop and qualify assessments that are intended for use across multiple drug development programs. So we

Page 33

work with many stakeholders in this pathway, including consortia, patient groups, individual academic investigators, and drug developers within the program to help develop and qualify publicly available outcome assessment tools.

We do have a guidance that describes the drug development tool qualification process. Outcome assessments that are used in clinical trials are not required to be qualified through this program, but we do believe that when assessments are developed in consultation with CDER, and then ultimately qualified through this program, this will help to encourage drug companies to pursue drug development in these areas because the companies can be confident that FDA agrees with the content and the measurement properties of a tool, thus lowering their risk.

So here's the link to our website that I mentioned a couple of times. I really do encourage you to take a look. And with that, I thank you, and I look forward to the discussion.

(Applause.)

Page 34

DR. STRAIN: Do you want to take one or two -- we could take one or two questions now, though we'll have question time as well after this pair of talks.

Let me go back actually, if I can then. First of all, it was a great talk. Thank you. Very methodical, and I liked how it laid everything out. Let me pose the question, but let you contemplate it while Dr. Kampman speaks.

What I'll be curious to hear about is your thoughts about how you take this structure and look at stimulant clinical trials, and what we need to do to move forward in that context, especially given the reliance and the addictions field on surrogate measures, as outcome measures so often for clinical trials. So maybe I'll plant that seed of a question.

DR. SILVERMAN: I have a question there, too, I could add. I wonder --

DR. STRAIN: Can you state your name?

DR. SILVERMAN: Oh, I'm Ken Silverman.

DR. STRAIN: For the recording.

Page 35

DR. SILVERMAN: Okay. I do research with incentives. I'm not attentive to the FDA concerns as I probably should be. But I've always considered the objective measures of a urinalysis to be the best outcome measure for our trials. But the FDA apparently, it sounds like, considers them less -- I don't know -- you need stronger evidence if you use those surrogate measures than if you used some self-report.

Could you just comment on that?

DR. SLAGLE: Yes. Well, I think there are trade-offs. So the objective measures are easier to measure in some cases, and they're "objective," so you feel like you can trust the results. Some people are a little less comfortable with the subjective measures; for example, patient reported outcomes.

But within our regulatory context, we are looking -- we have to identify treatment benefit, which is how patients, feel, function, or survive. And these surrogate measures, while sometimes easier to measure, don't always directly translate

Page 36

into how patients or function in daily life, or survive. So we need evidence that shows that this objective thing that you're measuring actually means something to the patient in their daily life.

So I wouldn't say that FDA doesn't -- we use surrogates in multiple conditions because there's existing evidence that tells us that this actually means something; it's meaningful to patients.

Does that help answer the question?

DR. SILVERMAN: Yes. Not sure I like the answer.

(Laughter.)

DR. SLAGLE: Well, just because you can impact some biologic process doesn't necessarily mean that it's doing anything beneficial to the patient. And so that's what we have to guard against, that we're not just changing something that we can measure without actually doing anything beneficial for the patient.

DR. STRAIN: So let me introduce Dr. Kyle Kampman, who's going to be speaking now on experience in developing a tool using the CSSA as a

Page 37

model. Dr. Kampman?

Presentation - Kyle Kampman

DR. KAMPMAN: Good morning. I'm going to talk to you today a CSSA, the Cocaine Selective Severity Assessment, which is a measure of cocaine withdrawal. Now, the purpose of my talk, as I understand it, is to use the CSSA as an example of instrument development. I do not believe that the CSSA is the ideal instrument to measure efficacy in cocaine pharmacotherapy trials. I just wanted to give that disclaimer. But what it does is it measures clinical phenomenon that seems to have implications in treatment outcome in outpatient cocaine-dependent patients, cocaine-dependent patients undergoing outpatient treatment.

So that's what it does. So what we're going to do for the next 20 minutes is talk a little bit about how this instrument came about. So what we're going to do is I'm going to introduce you to the CSSA, tell you where it came from, why we developed it, talk about what it measures, go through some of the basic reliability and validity

Page 38

testing that we did with it initially, and then talk about the subsequent testing we've done with it and the usefulness that we found for it in our clinical trials.

The CSSA, you've got a copy with you, in front of you, at your desktop, and it's a measure of cocaine withdrawal. It has 18 items, each one measured on a zero to 7 scale. And the signs and symptoms measured include things like appetite and sleep changes, cocaine craving, depressed mood, anxiety, irritability, lethargy, inattention, paranoia, and heart rate changes.

So if you do the math and you consider that two scales are mutually exclusive, your maximum score is 112. People generally don't score that high. On the whole, you get average scores in most of your treatment trials somewhere between 25 and 35. That's the scale.

So where did it come from? This scale was written by Joe Volpicelli at the Penn Center for the Study of Addictions. And he was my mentor, and I took over the testing of it shortly after he

Page 39

wrote it. And it's modeled on the Selective Severity Assessment, which is a tool that measures alcohol withdrawal symptoms. It's an old tool. It was first published in 1973. There are a couple versions, 30-item version, 11-item version.

Joe loved this tool. And he basically modeled it -- and those of you who are familiar with it will recognize the roots of the CSSA in the ASSA.

Anyway, Joe had this tool, and what he basically did, to begin with, is to comb the literature and look for symptoms of cocaine withdrawal as were reported by other investigators. And then he supplemented that with the signs and symptoms that the patients were reporting who came to us at Penn.

Joe was a pioneer in the outpatient detoxification of alcoholics, and that's why he was in love with these scales. And he believed that cocaine-dependent patients, like alcoholics, were negatively reinforced to continue their cocaine use by the withdrawal symptoms. So his ideal was to

Page 40

develop a medication -- first of all, to identify patients who had these severe cocaine withdrawal symptoms and identify a medication that could produce the symptoms and help them do better in treatment, or at the very least, identify a subgroup of patients who may have needed some more supportive psychotherapy or a more intense environment to get themselves clean. So that was our intention in the development of it.

Just to refresh the memories of some of you here and to introduce you to something maybe many of the younger people don't understand, during the late 1980s, cocaine withdrawal was kind of a big deal. It was very interesting to clinicians. It was a time during which cocaine-dependent patients were flooding our emergency rooms and our treatment centers.

They came with a constellation of symptoms that were really quite similar. They were usually depressed, sometimes very depressed, sometimes suicidal, often psychotic, primarily paranoid delusions. They had appetite changes. They had

Page 41

sleep changes. They were irritable and restless. And they were a mess.

Clinicians at the time tried to get these symptoms together and form clinical syndromes. Probably the most famous of these you'll see up here is the Gawin and Kleber Three Stage Withdrawal Syndrome, with the crash, the withdrawal, and the extinction phase. So people tried to put all these things together.

Other investigators, Bill Weddington, Sally Satel, and myself, really couldn't find a three-stage withdrawal syndrome. Basically, what we found is that patients came in with maximum withdrawal symptoms at the beginning of treatment, and they linearly declined as they got abstinent over the course of about 11 to 14 days.

Cocaine withdrawal is not medically significant. No one's going to get DT's or seizures from this. What really is important about cocaine withdrawal is the effect that it has on treatment outcome. Patients with more severe cocaine withdrawal symptoms just don't do well in

Page 42

treatment. They don't get clean, and more often than not, they drop out. So that's why we got interested in studying this particular phenomenon.

So that leads you to wonder, why do patients with more severe cocaine withdrawals do so crummy in treatment? Well, clearly, they could just simply be treating their withdrawal symptoms by returning to cocaine use. But we thought it might be more than that. We thought that these cocaine withdrawal symptoms were the clinical manifestation of the underlying neurobiologic changes that were occurring in the brain as part of the addictive process.

Part of what made us believe that were these studies that were starting to come out, showing that cocaine withdrawal symptoms actually increased the high that patients experienced from experimentally administered cocaine doses. And the first of these trials just showed the isolated symptoms -- for instance, depression or irritability -- in patients actually predicted a better high or an increase in the subjective

Page 43

effects of experimentally administered cocaine.

Of course, these are two biggies in cocaine withdrawal symptomatology. It's not the entire syndrome, but it's the beginning. So Mehmet Sofuoglu went ahead and took it a step further and showed that patients who actually met criteria for DSM-IV cocaine withdrawal syndrome actually had more subjective effects from experimentally administered cocaine.

In this case, he took 34 cocaine-dependent patients who had DSM-IV cocaine withdrawal, and he compared them to 10 cocaine-dependent patients who did not meet criteria for withdrawal, and he gave them cocaine. And among the other outcomes recorded were high and the effects of the last dose. And he measured that over 15 minutes, and what he found was patients who made criteria for cocaine-withdrawal syndrome reported a greater high from cocaine and reported a greater effect of their last dose of cocaine.

So maybe patients with cocaine withdrawal syndrome simply get a better high from cocaine, and

Page 44

that's part of the reason why these patients are so difficult to treat.

The human laboratory data combined with our clinical experience, really led us to believe that cocaine withdrawal was the tip of the iceberg, sort of the clinical manifestation of these underlying brain changes associated with cocaine dependence that included things like craving and hedonic dysregulation. And that is really why we became interested in measuring cocaine withdrawal and trying to -- or developing this scale.

So that's why we did it, and this is what it turned out to be. This is the CSSA. As I said, it's sitting on a desk in front of you. It's got the 18 items. It's an interviewer administer, which may be a weakness. We could actually easily convert this to a self-report scale, which would probably make it a whole lot more useable, but we haven't gotten around to doing that. We haven't done the reliability of the validity testing of the self-report.

So what remains, a clinician administered

Page 45

instrument, two craving scales, intensity and frequency. And mainly when we use it as a predictor of outcome, we're looking at a total score as one score, although we have looked at individual items, and interestingly, the two -- well maybe not so much. It would be predicted that the two craving scales would be predictive of outcomes. But what we actually found was bradycardia was almost as predictive as craving or the whole scale as a predictor of outcome. Think about that.

I am not a psychometrician. I am a psychiatrist, trained clinically. And Joe Volpicelli, when I was a young fellow, took this instrument, threw it on my lap, and said, "Kyle, go do initial reliability and validity test." And I said, "Joe, what's that?" And Joe said, "Go see Arthur Alterman," and so that's what I did. Arthur Alterman is a brilliant psychometrician at our center, and he taught me whatever I know about psychometrics. And I'm very grateful to his lessons, which were sometimes harsh, but still

Page 46

very, very useful. Those of you who know Arthur will understand that.

So what we did was, very simply, we did test/retest reliability. We did interrater reliability. We measured internal consistency, and concurrent validity and predictive validity.

Now, this I consider the boring stuff, and I'm not going to talk about it today, aside from saying that the instrument has really good test/retest reliability and very good interrater reliability. Its internal consistency is acceptable. If you do factor analysis on the instrument, it comes into two factors. Craving is a factor in and of itself, and everything else settles out as another factor. But there is pretty good item total correlations. But I'm going to talk a little bit more about the interesting stuff to me, which is the concurrent validity and the predictive validity.

Concurrent validity. The individual items on the CSSA that have the highest scores just happen to be those items that are part of the

Page 47

DSM-IV criteria for cocaine, which are all depressed mood, lethargy, increased appetite, increased sleep, and irritability. So that's pretty good.

We also wanted to measure whether or not our instrument coincided with other instruments of addiction severity. So one would guess that if one has more severe cocaine dependence, then one would have worse cocaine withdrawal symptoms.

So we looked at the ASI, and we found that we got very good individual item correlations with ASI severity measures. So for instance, scores on the CSSA correlated very well with days of cocaine use in the past 30 from the ASI, longer lifetime histories of cocaine use, and higher ASI severity scores for drug problems. So the patients who had more severe cocaine dependence tend to score higher on the scale.

We found that it was specific to cocaine withdrawal. It wasn't just some non-specific measure of addicts coming in for treatment being miserable. And we demonstrated that by showing the

Page 48

patients who came in with cocaine dependence and were newly abstinent from cocaine or were cocaine- and alcohol-dependent patients newly abstinent from cocaine scored higher on the CSSA than our plain alcohol-dependent patients. So it seemed to be specific to cocaine.

Finally, if this is in fact a scale that measures withdrawal, one would expect the scale to improve as patients became abstinent. And that's in fact what you see. And as I mentioned before, it's sort of a linear decline over time, and it takes about 11 to 14 days for patients to come down to baseline on their CSSA. And what's not shown and what I do have data to show is that patients who don't get abstinent don't have declines in their CSSA with repeated measures. So it does seem to behave like it's measuring cocaine withdrawal.

The most interesting thing about cocaine withdrawal to us, when we first started looking at it, was its ability to predict treatment outcome. Initially, we measured treatment outcome in three pretty different settings. One was a cocaine

Page 49

psychotherapy trial conducted at the University of Pennsylvania, and the second is from the IOP, or basic outpatient treatment, at the Philadelphia VA. And then the third context we studied in were medication trials being conducted at the University of Pennsylvania.

This is data from the first psychotherapy trial, was 87 patients. And what we measured here was who completed 30 days of treatment looking at patients with high and low CSSA. So in this case, the cut-off for this was 24, so scores above 24 were high; scores below were low. And again, patients with low scores on the CSSA were much more likely to complete the first 30 days of treatment. So it seemed to have a nice predictive validity in that.

Now, when we took it down to the Day Hospital, we ran into a little bit of a glitch. It turns out that if you have very severe Access 1 -- I guess we don't say Access 1 anymore. If you have really severe psychiatric severity, you tend to have really high scores on the CSSA, and

Page 50

they tend not to get a whole lot better. So if you've got a bad schizophrenic, a bad bipolar, a bad patient with PTSD, it may reduce the predictive validity. And that's what we found in this particular trial.

So to get this really pretty outcome of prediction of completion of 30 days of treatment at the Day Hospital, we had to exclude patients with severe psychiatric illness, which really doesn't make a lot of difference to us in our clinical trials, since we routinely exclude those patients anyway, but it does have implications for use in general.

So that's one of the weaknesses of the CSSA is that it may not necessarily be predictive in patients with more severe co-morbid psychiatric illness. And the cut-off on this was a little bit higher. It was 37, was what we defined. And you're going to see a pattern here. Again, scores in the upper 20's to 30's tend to be the high scores that predict. And if you want to look at just the percentage of subjects, it's generally the

Page 51

upper one-third. Patients who have scored in the upper one-third on the CSSA tend to be the ones that have more problems in treatment.

Finally, medication trials. This was a bunch of open trials, I think 3 or 4 open trials, 76 patients who participated in those. And what we were measuring in this case is 3 weeks of continuance abstinence at any point during those brief trials. And again, we see a huge difference. Cut-off score in this case is 21, but again, it's just pretty consistent that if you have low scores on the CSSA, you're much more likely to do well.

What are the advantages of being at a center like the University of Pennsylvania, aside from having Chuck O'Brien and Arthur Alterman and Tom McClellan to help you, is that you can take your instrument and you can put it into every single clinical trial that gets done at the center, which I did. So every -- with the exception of some of the multicenter trials -- it won't allow me to do that -- I will stick the CSSA into every single cocaine trial that's been ongoing for the past

Page 52

20 years. That's a lot of trials.

What we did is we decided to go ahead and put the data together from as many of these trials as we could and identify predictors in cocaine dependence treatment. What it turned out to be were 7 fairly large clinical trials, all of which were about 7 to 12 weeks in duration. We excluded the alcohol, the comorbid alcohol-dependent patients. Had I included them, I could have probably the tripled the end, but we looked only at the cocaine-dependent patients.

We had outcome measures. And the beautiful thins is we're at 10, so everybody gets the ASI as well as the CSSA. We got urines 2 or 3 times weekly, and of course we get timeline follow-back self-report on everybody that comes through. So for predictor variables for outcome, we had, as I said, the ASI urine drug screens, and we had the CSSA on all these folks.

For outcomes, success and treatment, we arbitrarily selected 3, 3 weeks of continuance abstinence at any point during the trial. This is

Page 53

urine's negative, none missing. So it was a hard abstinence measure.

We also looked at 50 percent reduction in ASI composite drug scores. Chuck likes that one, so we used that. That seems to be a reasonable measure of improvement. And then finally, we also looked at no self-reported cocaine use during the last 4 weeks of the trial, just to get rid of the whole idea of urines; let's just see what the folks tell us. And we used that as outcome measure.

What we ended up with was our usual sample of folks. Now this trial was actually done a few -- the trials were done a few years ago, so the age actually has increased. Our average age right now runs about 49 to 50 years old among our cocaine-dependent patients. It's an aging population. But we gathered up about 402 of these guys, and as usual, they are primarily African American men who smoked crack cocaine. Days of cocaine use, 13. That's pretty standard, about $600 of cocaine in the month prior with an ASI composite drug score of .23 and a baseline CSSA of

Page 54

27. So that's sort of average.

We took all of our predictors together, and if you look at the 3 outcomes here, the most consistent predictors of outcome when we threw them into the regression were initial urine drug screen and initial CSSA scores, both of which came out to be significant in the regression model independently. So they don't overlap. They're different. They measure different things. But these are the two things that best predict outcomes in our clinical trials.

What about predicting medication response? Can we use baseline CSSA scores to identify patients who might respond better to medications? And we were excited about this one in the beginning. This is amantadine. This is a medicine that we thought was going to treat cocaine withdrawal symptoms, and it was the first trial I ran at Penn. And the trial as a whole was negative, but when we went back and we looked, and we separated out patients with high and low CSSA scores, again, using a two-thirds/one-third

Page 55

cut-off, the patients with high scores in the CSSA seemed to have a wonderful response to amantadine in the number of clean urines. So that got us really excited.

Unfortunately, this is a really small sample. This is a 60-patient study on the top there, so there are 20 people in this group. So when we went back to replicate it, it didn't replicate, and we have of course long since given up on amantadine. But I just though I'd throw that in for historical purposes.

Propranolol is another drug that I studied early on in my career, which seemed to have a differential response to patients with more severe withdrawal symptoms, which again sort of made sense to us. And this is, you again take the patients with the top one-third on their baseline CSSA, which I think was a score of 28 in this case, and you just look at treatment retention or urinary benzoylecgonine levels. This is actually quantitative urinary benzoylecgonine levels, which we actually did measure and use as an outcome, in

Page 56

the past. But in any event, if you got propranolol, you're more likely to be retained in treatment, and you had lower overall urinary benzoylecgonine levels. So that's great.

This we partially replicated in the subsequent trial of propranolol in that we found that propranolol among patients who actually took their propranolol and had high CSSA scores actually did look a little bit better than placebo, nothing to write home about, and I don't think we're going to pursue propranolol as a medication going forward, but it was replicated.

Now, the interesting thing about propranolol is this is the one and only medicine that I have ever studied that showed a differential change in CSSA scores during the trial compared to placebo. So propranolol actually lowered CSSA scores during the trial, and it's the only one of the medications I've studied so far that actually does that.

Now, most recently, we looked at topiramate, and we've done two trials of topiramate at our center. Bankole Johnson has a positive trial

Page 57

there. So actually there's three positive trials of topiramate out there in the record, and we have two of them. And this is our data from our second trial, which was cocaine- and alcohol-dependent patients, 177 subjects, 300 milligrams of topiramate. And in the main trial, we found that topiramate was associated with greater end-of-study abstinence compared to placebo, so that was good.

So we went back and we looked at baseline predictors of topiramate success, and this is the only thing that we found. Patients with high scores on the CSSA did better with topiramate. And what we did was we divided the sample into tertiles, so we've got your low CSSA scores, your medium scores, and your high scores. And we only saw a topiramate effect in patients with high scores, which was CSSA above 18.

All right. That's the good part. Now, although we have really nice predictive validity in initial CSSA scores, what the CSSA falls down in is in outcome measures. The fact that we haven't yet found a medication that actually predicts a

Page 58

differential response to placebo in reducing these scores. So what you see in clinical trials is that the mean CSSA scores tend to decline over time in all our trials. That's pretty much universal.

This is actually probably not even an exhaustive list, but these are just the ones I thought of when I was putting together and checked. None of these medications actually showed a difference in CSSA over time, so that was kind of disappointing, with the exceptions I said of propranolol. So as an outcome measure, maybe not so good.

What do we know about the CSSA? CSSA is a good predictor of outcome in outpatient cocaine-dependent treatment. Aside from urine drug screens, it's probably the best predictor. So what do we use it for primarily? We use it as an earned variable to make sure that our poor prognosis patients are equally distributed between a placebo and the active group, and that's something that we do. We earn pretty much everybody in our cocaine trials on urine drug screen on the day of

Page 59

randomization and their CSSA score on the day of randomization.

The CSSA may identify subgroups that are responsive to particular medications. Propranolol is an example. Topiramate may be a better example. Then finally, no medication tested thus far has shown a differential response in reducing CSSA scores consistently. Propranolol did it once, but it didn't do it in the second trial.

Finally, I just want to make a plug for NIDA centers, wonderful things. Trying to do this kind of work outside of the context of a NIDA center would have been extremely difficult. Not having all those people funded in the same building with me to help me; not having the ability to insert my instrument into all the trials being done at the center, and not to have the funding and the ability to do all these secondary analyses and get all these patients published, again, it would be very, very difficult. So I'm extremely grateful to NIDA for allowing us to have the center for all the years that we've had it and continue to have it

Page 60

because it does support things like that.

With that, I'll stop.

(Applause.)

Q&A - Group Discussion

DR. STRAIN: Maybe we could take a couple minutes for questions.

DR. KAMPMAN: Sure.

DR. STRAIN: And I'm actually going to start. Is the CSSA copyrighted?

DR. KAMPMAN: No, in the public domain.

DR. STRAIN: Okay. It's an important point simply because as we move to electronic medical records, not just for clinical care, but for the interface of clinical care with research, we're getting push back for copyrighted instruments like the CINA, for example, with opioid clinical trials, which is copyrighted. So then the lawyers descend upon us.

Do we have lawyers in the room?

(Laughter.)

DR. STRAIN: So the lawyers descend upon us and tell us that essentially we can't use those

Page 61

copyrighted instruments. So I'm very glad to hear the CSSA is not copyrighted, and I'd encourage you to continue in that venue.

DR. KAMPMAN: We want it used.

DR. STRAIN: Yeah. Good. I had also just one other question before we go to the discussant. What do you think is the next step in the CSSA development?

DR. KAMPMAN: The CSSA was sort of put out there, and I got interested in finding medications and accepted it for what it was. The instrument has never been modified, improved in any way. This is basically what Joe Volpicelli gave me back in 1992. We could fix up some of the items. It certainly would be better as a self-report. That's one of the biggest complaints I get from people. Why isn't this self-report? This doesn't make any sense. You don't need to have a research tech. Ask these people these questions. So that would be one of the things we would do.

DR. STRAIN: Thanks.

Other questions? Yes, Kathy?

Page 62

DR. CARROLL: Thinking about how it might be developed as an outcome measure, have you done anything looking at change across time? And is there any evidence that if people kind of go from that high level above 21 to somewhere below it, some set point, that that predicts how they do in the long-term?

DR. KAMPMAN: No. We've looked at it mainly as a predictor --

DR. CARROLL: As a predictor, but not as a --

DR. KAMPMAN: Not as over time. What we've looked at more as a predictor is change in urine drug screens, abstinence during the first two weeks. Cocaine trials are like cigarette trials. The best predictor of who's going to do well in your trial is what happens during the first two weeks, but we haven't done that with the CSSA. The CSSA just kind of goes down over time.

Yeah, George?

DR. STRAIN: Ivan was waiting, and then George.

Page 63

DR. KAMPMAN: Oh, I'm sorry.

DR. MONTOYA: I just have a question. Maybe I missed it.

DR. STRAIN: Identify yourself for the recording.

DR. MONTOYA: Ivan Montoya from NIDA. What's the sensitivity and specificity of the --

DR. KAMPMAN: Sensitivity for the CSSA is really, really good. It approaches like 80-90 percent. The specificity stinks, mainly because -- and that is if you're predicting poor outcome. Because so many people do poorly in our cocaine trials, the sensitivity is very high and the specificity is pretty low. So it identifies people who are going to do crappy really well. It doesn't necessarily identify people who are going to do well.

DR. STRAIN: George?

DR. WOODY: George Woody from Penn. It's a question for the FDA. Many of the drugs are illegal, and arrests in illegal activities is one of the ASI factors. How does that count in our

Page 64

outcome measures? Because you can show decreases in arrests and crime and various things in some studies.

DR. HERTZ: This is Sharon Hertz. I'm with the division. And I'm not sure I understand your question fully because when you're assessing a drug to treat addiction, what we want to know is that people stop using the drug. The fact that there is an activity of illegal behavior associated with drug use is clearly a bad effect that's associated with using drugs, but -- are you trying to ask if it would be a suitable surrogate?

DR. WOODY: That's just the question, is that in the various outcomes, that wasn't mentioned, but it is part of the ASI. And of course, that was studied pretty heavily with methadone. That was one of the things with methadone, that it reduced crime and police -- so I was just curious.

DR. HERTZ: Right. So it's an important element. It could be an interesting secondary measure. I'm not sure that it would have

Page 65

necessarily labeling value.
DR. WINCHELL: I'm Celia Winchell, also with the FDA. I can't speak into the microphone and look at Dr. Woody at the same time, so I'm going to apologize to Dr. Woody and speak into the microphone.
DR. WOODY: That's fine.
DR. WINCHELL: There are a lot of adverse consequences of being involved, addicted to various substances, and criminality is certainly one of them. But we think it's probably unrealistic to run a clinical trial long enough that you could directly measure impacts such as criminality, although a broad basket of adverse consequences of drug addiction, that's kind of the direction that we looked at when we were exploring what patterns of alcohol use behavior, other than ceasing alcohol use altogether, could translate to clinical benefit for patients.
There were batteries of a variety of alcohol associated consequences, which include -- I mean, legal problems such as driving while intoxicated.

Page 66

One doesn't so much get arrested for possession. And though that was part of the battery of consequences that was evaluated against various patterns of alcohol use, conceivably among the many bad things that could happen to a person as a result of his or her involvement with drugs, criminality could be one aspect that you would look at long-term, but it doesn't seem as a practical endpoint for a clinical trial.
DR. SLAGLE: This is Ashley Slagle from the FDA also. I think criminality is one of those things that is more distally related to the treatment and the condition, so it's a downstream effect. If you treat the -- if the patients become abstinent, because of that, criminality is reduced longer term like Celia said.
There are also other things that contribute to criminality, so psychosocial support and all of these other variables that contribute. So it's a little bit harder to interpret a change on criminality without also being able to understand that there's abstinence or that the drug use has

Page 67

been reduced.
Discussant - Eric Strain
DR. STRAIN: Let me interject here, and I want to switch hats and become Bob Dworkin for a couple of minutes and summarize the two talks because I think we are moving into some questions, and I've got questions as well, certainly, broader about it.
But before we do that, let me thank both the speakers, first of all. I thought they were great talks. I think they nicely bookend each other in the sense that Ashley's talk kind of giving us a real structure and a way to think about the methodology and about what are the parameters related to this topic. I'm looking forward to looking over your slides again. They were rich slides, I thought, that really helped to give a structure to this. So I greatly appreciated it.
One of the things that I noted from your talk, Ashley, that I think I'll probably come back to if I have time for questions, is this concept of how the patients feel, function, or survive. I

Page 68

think that's a very intriguing thing to consider when we're in the addictions field because it's different from what we usually think about. And this resonates also with Ken Silverman's comment about urine test results, which I think we'll probably come back to some more while we're here.
I greatly appreciated, as well, your points about -- all points, but about assessments being well defined and reliable. And I thought that nicely resonated with Kyle's talk, which I'll talk about in a moment, where he was looking at both the validity and reliability aspects of the CSSA.
I think that you also touched on the concept of what is a meaningful change on a measure, and meaningful change compared to risk. And risk didn't really -- you didn't really dive into risk a lot, and I think that may be an opportunity as well for us to think about and to have some discussion about because there are risks inherent in stimulant-use disorders. So change needs to be interpreted in the context of that risk, which is something to consider.

Page 69

I thought, as well, the discussion on -- you touched on benefits, both direct and indirect, and the outcome assessments. Both patient related, clinical related, and surrogate measures all were intriguing points and things that may lead to further discussion by us.

Your outline of how to assess an instrument led I think nicely to Kyle's talk about the CSSA, this instrument that's been around -- I didn't realize it's been around now for 22 or 23 years. It was fun an interesting to hear about the history of it as well. I didn't realize that Joe Volpicelli had originated it out of the ASSA.

Kyle was careful to caution us about what this instrument is, and that it's not -- in some ways, it sounds to me like -- and I don't mean this in a disrespectful way, but you kind of backed into this as something that perhaps took on more of a life of its own than you initially thought it might as an instrument although it has resulted in you getting multiple papers out of it, which is nota a bad thing if you're in an academic medical center.

Page 70

There was this intriguing point about its content, especially with respect to bradycardia. And I thought you were going to draw something out there about the propranolol effects and bradycardia perhaps, if there was something that resonated there, and maybe you'll want to comment on that at some point.

There also, though, is this feeling with the -- so I appreciated that you went through and discussed a little bit about the steps that you went through in test/retest, and interrater, internal reliability, internal consistency, and then concurrent and predictive validity.

I think the important thing to recognize out of this discussion or this presentation is that this has not been developed as an outcome measure. It's really not been a robust outcome measure, but it has shown some value in being predictive, which is interesting and intriguing. And if anything, I think in your free time, which I'm sure you have lots of, it would be interesting to think about how to take the instrument and make a next generation

Page 71

of it, it seems to me, to try to further refine it and see if there could be further value to it in some way, perhaps either as a more robust predictor with a specific score rather than a somewhat floating cut-off score that seems to vary across clinical trials, or with a further refinement of the content, or as a self-report measure, of course, which would be valuable.

Those are my initial thoughts about these two talks. Oh. And I just want to acknowledge that Kyle in his talk also talked about cocaine withdrawal and the -- I appreciated the historical context of that, that three-stage model that Gawin and Kleber showed, which I don't think anybody ever was able to substantiate, but it got them a paper in the Archives of General Psychiatry at the time. But there was this constellation of patients with cocaine withdrawal, certainly, that was interesting.

The only other thing that I thought and we may want to come back to is I was intrigued that the high CSSA scores were associated with better

Page 72

outcomes in your amantadine clinical trial, if I got that right. And that seemed a little paradoxical to me in that -- I think I've got this right -- in the modafinil studies, lower -- isn't it the case that the positive signals in the modafinil clinical trials have been associated with lower rates of methamphetamine use?

DR. KAMPMAN: Yes. We could not get the CSSA to predict outcomes in modafinil, which was very disappointing to us. We really thought it would, but it did not.

DR. STRAIN: Yes. So there's something a little paradoxical there.

Let me stop there, and maybe we can open it up for discussion because I think there's a lot of potential areas to discuss. So Sharon and then Steven.

DR. HERTZ: Thanks. This is Sharon Hertz with FDA. So I'm interested in a couple of points about whether or not the CSSA can be used as an outcome measure. It seems to me that because we don't really have treatments that we've identified

Page 73

as being particularly useful, that I'm not sure the failure of CSSA in those studies necessarily means it's not going to be a good measure once we actually find something that works. So I'm not sure that I have any concern about that lack of a good outcome yet since the drugs have failed us and perhaps not the instrument.

But my question is, the predictive measure of the CSSA seems very useful when you're initially assessing somebody in a state of withdrawal. And I think that -- my question is, how confident are we that drugs that are used to treat the addiction -- I'm having trouble quite making this connection, and that's the question I have.

How would one connect the withdrawal period and the long-term success for avoiding relapse?

DR. KAMPMAN: We conceive medications now, at least when we're studying them, as medications that are essentially helpful for abstinence initiation versus those that are effective for relapse prevention, and the two may be completely different.

Page 74

Now, just treating withdrawal we know in alcohol doesn't help long-term because people just relapse. So the only advantage of treating cocaine withdrawal symptoms, if you're really concerned about that, is to stop this sudden drop out in treatment. And that's really what we thought about in the beginning. Even when we were doing clinical trials for cocaine dependence, we would get 30 or 40 percent of our patients to be retained in an 8-week trial. That was success. That was crazy.

So we were kind of hoping that we would treat the withdrawal symptoms, they'd stay in treatment, and then we would find another medicine that might help them, or psychosocial treatment engagement would help them down the road. This is just to get people engaged and get them initially abstinent.

DR. STRAIN: Steve?

DR. SPARENBORG: Kyle, thanks much for the good explanation of the CSSA. Ashley's presentation about the underlying meaning, when she talked about this underlying meaning of what you're

Page 75

looking at, this particular questionnaire or whatever, does the CSSA correlate in any way to measured or self-reported 30-day prior to baseline cocaine use, severity? And if it does, what does that mean? And if it doesn't, what does that mean?

DR. KAMPMAN: Yes, it does. That's one of the ASI variables that correlates pretty well with baseline CSSA scores. It's 30 days -- the number of days of cocaine use in the prior 30 days. So we always attribute that to patients who use more severely should have more severe withdrawal symptoms. So that does correlate, and that we agree with.

Could you use that measure as a surrogate of CSSA scores? Yes. They don't work as well. We've put those into the regression models, and days of cocaine use will stay for a while but will fall out, whereas CSSA scores will remain the same.

Interesting that you brought that up because another story of topiramate -- I'm sorry. I'm stuck on topiramate. But another study of topiramate done by my partners at Columbia, who

Page 76

we're collaborating with in a replication trial, found that patients who used more frequently in the 30 days prior to entering a trial actually were the ones that responded to the combination of topiramate and Adderall, which kind of harkens back to my finding of high CSSA scores predicting a good outcome with topiramate. Just thought I'd throw that out there.

DR. SPARENBORG: Just one last comment. I think it's important for us to try to strengthen the relationship between self-report and any more objective measure of drug use, and I think this is very important for us to consider during this meeting.

DR. KAMPMAN: Thanks.

DR. STRAIN: David?

DR. McCANN: Dave McCann from NIDA. First I want to echo the thanks on both of the presentations. They were wonderful. Part of what goes through my mind is what we might be talking about at the next meeting of this group that might be on the next agenda. And before I forget, I just

Page 77

want to mention that stratification factors for clinical trial design is something that we should definitely talk about.

I hadn't considered the CSSA as something we should look at, but it certainly sounds good. The problem is, when we sketch out a protocol and run it by our statisticians, they don't like to have too many stratification factors. We have to start out asking for three so they will compromise and take two. We've looked at things like alcohol dependence, whether the last urine during screening is clean or dirty. CSSA would be a third one we'd add on there. It's tough to get those statisticians to work with three of them.

I don't want to get into a long discussion here, but at a future meeting, I think that this is something we need to think about. If we're limited to two, what are the best two? The other thing would be how do we convince statisticians to use three because -- to me, if you see three factors that are highly predictive of outcome, is it better to totally ignore one so that you can have a great

Page 78

balance of the other two, or to try all three and come close with all three? That's something we need to talk about.

But getting back to the CSSA --

DR. STRAIN: So now we've dissed lawyers and statisticians.

(Laughter.)

DR. McCANN: So a question about the CSSA; there is a question in here. To what extent is this really withdrawal? If you bring somebody inpatient with high CSSA scores, and you give them cocaine, now they have almost no score?

DR. KAMPMAN: Have I ever done an inpatient withdrawal test with the CSSA? No. What has been the findings of cocaine withdrawal among patients who are brought in to inpatient settings, it goes away, pretty much, and you don't get a lot of symptoms inpatient.

In fact, when I wrote the initial paper, I had to take every reference to withdrawal out of it. If you look at my initial reliability of validity testing, there is no cocaine withdrawal

Page 79

written there because I couldn't get it pass the editors because at the time, no one believed that cocaine withdrawal existed. And that was mainly Weddington and Satel's two papers.

I think that if you put patients inpatient, you reduce their craving significantly, and that is a large part of the CSSA, which is why you get I think much lower scores. But you still will get some withdrawal that may not be as pronounced as you get on an outpatient.

DR. STRAIN: Yes?

DR. FALK: Dan Falk, NIAAA. From the CSSA presentation, I got that it's not a great outcome measure, but it could be a good moderator of treatment effect. I guess my question is maybe for the FDA, Ashley maybe. I don't know if anyone would necessarily want to go for the CSSA as a PRO patient reported outcome, but maybe as a moderator they'd want to put it in a phase 3 trial perhaps to say that, well, we think that the treatment effect might vary as a function of their CSSA score.

Is there anything special that needs to be

Page 80

done from a PRO perspective or from maybe an FDA perspective on creating subgroups or validating this instrument for moderator purposes?

DR. SLAGLE: My group typically focuses on outcome assessments, but I think that gets back to the question that David had about the stratification and whether there's a benefit to really doing that with other variables that you can use. I'm not sure if my clinical colleagues have any additional thoughts on this, using the CSSA as a moderator in trials rather than an outcome.

DR. WINCHELL: So if I understand correctly, Dan, your question was should it turn out that this tool or a tool like this was actually very helpful to distinguish between patients who really would benefit from medication treatment and potentially patients who wouldn't, and therefore could be used maybe in an enrichment strategy, would any further work on that tool need to be done for it to be appropriate for use in a registration trial. So that we would say in the label, this medication is beneficial for the subset of the population who

Page 81

have this diagnosis and meet this other criterion.
I would think it would have to be at least something that we had taken a look at and thought -- I mean, it couldn't be something completely not vetted at all. I'm not sure what the process would be for that, but it would be incorporated as almost a diagnostic measure, the person has a condition that will be responsive to this medication. And maybe we kind of -- we're used to looking at that type of thing in the context of an NDA. We've got things like that.
DR. FALK: Thanks.
DR. HERTZ: This is Sharon Hertz. I think that the work that was maybe alluded to that there's already been some interest in having this developed as a PRO, so just whatever it would take to transition the instrument, whatever the basic work would be, could be that type of psychometric -- and we use the term "psychometric conversion" because it sounds good in my head. I don't know if it's a term.
(Laughter.)

Page 82

DR. HERTZ: But whatever that is, yes. The value of something as a tool to help select patients is whether it's available to clinicians, is it something they're likely and able to use. So the easier it is to have someone use the tool, particularly if it becomes appropriately transitioned to a PRO, it certainly could have some promise there.
DR. STRAIN: So just to clarify -- because one of the outcomes of this meeting is what should we -- are there things that the field should work on. So Dan, I appreciate your question in that context because this could at least a beta version of something that could be developed, is what I hear, that may be of value either as a stratification or as a moderator, although it may not be this particular instrument, I guess.
I'm thinking because it's an interviewer rated instrument, it's not a -- quite honestly -- it's 18 items I think.
DR. KAMPMAN: Yes.
DR. STRAIN: Yes. An 18-item

Page 83

interviewer-based instrument is not something that may be used clinically as much a self-report instrument.
Somebody in the back was -- go on, please, Laurie.
MS. BURKE: I think that --
DR. STRAIN: You want to identify yourself for the --
MS. BURKE: I'm Laurie Burke, and I am -- what am I today? University of Maryland or I also am LORA Group, whichever. I am wanting to perhaps ask -- keep a focus on what exactly -- the CSSA measures and the write-ups as it measures cocaine withdrawal signs and symptoms.
I think that in the discussion, it would be really useful to make sure that we keep our eye on that. I mean, there was discussion about how it doesn't really measure the long-term abstinence because it doesn't measure that. It measures signs and symptoms. But how is that related?
I think we need to go back to Ashley's presentation of exactly what are we trying to make

Page 84

this measure do in a clinical trial. It's hard to identify the value of this measure unless we know exactly what we want to be able to conclude at the end of the day by the score that's generated.
If it's cocaine -- and I want to compare this with depression. In psychiatry, there are depression signs and symptoms that the clinician observes and comes up with a conclusion of the diagnosis of depression. This seems to me very similar, where there is a definition of cocaine addiction -- or what is the DSM diagnostic category?
DR. STRAIN: Use disorder.
MS. BURKE: Use disorder. And within that, you identify things, the signs and symptoms, which are also included in the CSSA.
In the development of this instrument or of the alcohol instrument that this is derived from, was there any actual interview of patients to make sure that they are agreeing that the signs and symptoms here are inclusive of all the most important things in that course that cocaine

Page 85

addicts experience?
I think that would be an important first step, to first making sure that you have not just the diagnostics of signs and symptoms, but the things that actually are experienced by the patient, which is getting to the important thing about what do patients actually experience, feel and function in their daily life.
My question is, has that been done to any extent? And if not, how could that be a next step for the development of this for an outcome assessment?
DR. KAMPMAN: We did not do focus groups, but what we had was a busy clinic with nurse practitioners, cocaine-dependent patients, and Joe Volpicelli. Although some of the items in the instrument came from literature review, a great number of them came from what our patients were reporting to us. So, no. We did not do focus groups.
DR. STRAIN: Connie, and then I'm going to ask a question, then Ashley.

Page 86

DR. WEISNER: Connie Weisner from Kaiser Permanente and UCSF. I think that this moves a little bit more to the outcome part, but working -- doing research in a health plan that has a very -- we find that our cocaine patients are very heterogeneous. We use the ASI in our instruments. And that question for each of the domains that says how important to you is treatment for your drug problem, how important is it for your legal problem, how important is it for your employment problem, those people, those cocaine-dependent people, or other stimulant-dependent people, who are there on a legal referral say it's extremely important for their legal problem. Treatment is not important -- very important for their cocaine problem.
If they're there on an employment mandate, it's the same. They're there to fix their -- it's extremely important to fix their employment problem, not extremely important -- so the job -- the way the clinicians see their job is to

Page 87

make this connection for people that if they stop using cocaine, their employment problem will get better and so forth.
I'll talk more about this tomorrow, but kind of teasing apart what patients want out of treatment and what stakeholders want out of treatment, that sort of sometimes disconnects.
DR. STRAIN: Thanks. Actually, my question is very congruent with that comment, so I really appreciate it. I want to go back to the FDA on this, and maybe I've just gotten hooked on this, but how patients feel, function, or survive, that statement. And I think that's a quote, right? Have I got that?
(Laughter.)
DR. STRAIN: And it's feel, function, "or" survive. It's not "and survive." So you'll accept feel, function, or survive. Am I being like a lawyer now?
(Laughter.)
FEMALE SPEAKER: Or a priest. I don't know.
DR. STRAIN: Yes. Bless you.

Page 88

So you didn't agree or disagree on the "or" question, but let me move on. It's an interesting thing. I think, in addictions, if we add a medication trial, and we had an outcome that produced improved functioning but didn't necessarily substantially impact drug use, would the FDA find that persuasive as an efficacious medication?
Let me further elaborate. I can see all the FDA people actually move towards their microphones.
DR. HERTZ: Actually, I would like to answer that by asking us to focus on other kinds of questions because we're really not here to ask about what is necessary to get a drug approved. We're here to find out how to evaluate this area. And we're really -- our participation is to sort of provide information, but I would like to not answer that question.
So what I would do is I would like to take that question and turn it back to all of you. Would you find a drug that improves function without affecting drug use to be something that the

Page 89

clinical group finds important and would be helpful to the community for whatever reason? And then that argument, if it was to be something used in drug development, could then be brought through the proper channels for discussion.

So it's not what's important to us in terms of these outcomes. It's what important -- and this is a pretty esteemed group of people within this field. So what's important to you, and then how can we assist through these mechanisms to make sure that the instruments are available to support that work, and how it then intersects with specific product development.

DR. STRAIN: And I fully appreciate and respect the answer. The reason I ask it, not as to -- I'm not trying to bait you into something or the FDA into that. What I'm trying to do is, I'm trying to figure out should we look at developing an instrument as an outcome measure that emphasizes function.

This goes back to Ken Silverman's question from like an hour or an hour and a half ago, which

Page 90

is we've been as a field obsessed with drug results, urine results, I think, as sort of our standard for an outcome measure. But it's interesting -- and let me riff on this for a moment.

It's interesting to think about something like a functional outcome because I found myself thinking, what if you take somebody with a chronic condition like rheumatoid arthritis and say, we're going to work on a medication that looks at functional outcome rather than cure to rheumatoid arthritis? That probably would be something that's got real value to that patient population. Well, it does. I mean, we've got medications that treat the symptoms and improve functioning without undressing the underlying disease state.

So I've not really thought about stimulant-use disorders as something where we should focus on outcome measures that look at function. I've always looked at it as let's focus upon the drug use itself.

I'm kind of thinking aloud and digesting

Page 91

this idea, but it was a really intriguing idea to think about it in other clinical conditions.

David?

DR. McCANN: Kelly can go ahead.

DR. DUNN: Just to echo, not dysfunction but also how patients feel.

DR. STRAIN: Yes, function, feel, or survive. Yes.

MALE SPEAKER: The question, function, feel, or survive. The question, function, feel, or survive. So let's say you could find a drug that would help with something like that, but did not decrease cocaine use. That's really the kind of question. I'm trying think what parallel there might be in another addiction disorder. How about naloxone? I mean, you don't use it to treat addiction, but it has great benefit for the patients.

It's worth looking at these issues. It may not be an addiction treatment. It may be something else that's beneficial for the patient.

DR. STRAIN: And I'm not -- just to be

Page 92

clear, I'm not advocating it as the only thing, but in sort of a menu of things that we should be thinking about in terms of outcome measures, the development of outcome measures. Should there be this slice that we should be thinking about, which may be disentangled from the drug use slice in some ways.

Laurie? Was somebody else before Laurie? No? Okay. Laurie.

MS. BURKE: Well, I think that this conversation -- and I'm not part of your expertise here in terms of treatment of addiction disorders. But I think that going back to what Ashley presented in that roadmap is going to be really useful to helping you sort out the answer to this conundrum that you're pointing out. Because what you measure and what you develop an outcome measure for is going to completely depend on all of that background information in column 1: what the patient population looks like; what they're seeking in terms of treatment; what the community is desiring in terms of their treatment; what other

Page 93

treatments are available; what they do when they don't have an successful treatment; the healthcare environment that exists; how they're identified; how they're self-identified.

I think having that to inform then your column 2 of how you're going to actually set up a trial, what are the objections of your trial, is it going to be just signs and symptoms; and it might be valuable to have a treatment that just gets them through that initial withdrawal signs and symptoms base. And then have an extension study to look at other outcomes.

The design of your study is going to determine what you really want to find out in the context of the treatment that you want to test. I look at all of the negative studies, and like Sharon said, you can take that as we have no treatments that are effective or you can take that as we don't have a measure to know whether there were little effects that were useful here. We don't know either way because the outcome measure that we're using is rather blunt. It's better for

Page 94

a diagnostic. It's not specific to the symptoms syndrome that the patients are experiencing.

There's one more thing I wanted to say; and the fact that it worked with patients who had very high scores at baseline. That also tells you that the sensitivity is not there -- I mean, we don't know. That's another research question, can the measure be better to measure change in those that have less severe symptoms at baseline?

Perhaps your research objective would be just to study those people with high symptoms at baseline and exclude others from your trial, just to start with, just to try to figure out the whole outcome versus the treatment effect issue, and the whole enrichment trial kind of ideal. I don't know if that would be acceptable to the division or not.

But I think that in order to answer the question about what should we do, proceed which direction, with this outcome measure, you really have to put together that background and that difficult thinking about what's your context of use, what's your clinical trial design, entry

Page 95

criteria, and also then design what the thing is that's useful to measure as opposed to just putting in the tools that are readily available.

I think that's the whole purpose of impact and action, is to get people thinking. So I don't know. What is the low-hanging fruit in terms of the context of use and that background information? And I'm sure this room has all the possibilities -- all of that information to able to put together some sort of a research agenda or a priority for the initial foray into developing a measure for a certain context of use and the thing that you want to measure in that context. And that would be my question. What is that, is the most important thing.

DR. STRAIN: We're running into our break time. And I think those are great questions. I think those are the meta-questions that we're struggling with here. Let me encourage people to continue the discussion while we break, but I do want to respect the break. I also want to thank our two speakers from this morning. And let's

Page 96

reconvene in here in about 10 or so minutes if we could. I think that's what we have. And Laurie will be certainly resuming those questions over the course of the day.

Thanks. Thank you guys.

(Whereupon, a recess was taken.)

DR. STRAIN: We're now going to hear from Raye Litten, who's going to provide us an overview to the experience and developing an outcome for alcohol studies. Without further ado, Raye.

Presentation - Raye Litten

DR. LITTEN: Thank you. And thank you very much, Eric. It's a pleasure for me to be here today. What I'm going to present is an exploration on our part here at NIAAA on endpoints, and really focusing more on pivotal clinical trials that treat alcohol use disorder.

First, I'd just like to mention key organizations that are really kind of dedicated to improving the methodology of clinical trials. And we're hoping by improving the methodology of clinical trials, we can increase our sensitivity to

Page 97

detect a difference in our medication during a clinical trial. I'll mention our NIAAA team, Dan Falk, Joanne Ferig, and Megan Ryan. And they oversee the human studies and medications development.

This is really a great team. They really work well together. Megan and Joanne are both here, as well as Dan who will be our discussant on this. I'll just mention the ACTIVE Group. This is another acronym. Don't even ask me what it stands for. I've forgotten at this point. Being in the government, we have so many acronyms. But what really gets confusing is when they use the same acronym for different organizations, then that really kind of gets confusing.

But this group has been meeting for years. It's kind of unique. It has the FDA present. At times we have the EMA, which is the European counterpart of the FDA, regulatory affairs. We have pharmaceutical companies, anywhere from six to eight, and some of these are European companies. We have academic researchers. Some of them are

Page 98

international researchers. We have NIAAA present. Dan and I primarily attend these meetings. We also have NIDA. Dave McCann is a regular participant, and also from the academic side, Chuck O'Brien comes to all these meetings also.

With that, I'd like to talk about first, FDA has put out a draft guidance. And I can tell you the field's very happy and excited about them putting this out. This is certainly the first one put out in the alcohol field, and maybe even the first one put out in the addiction field.

Basically, they have made recommendations to use two endpoints. You don't have to justify using these endpoints. The first one is, of course, the old gold standard percent subjects abstinent. That's usually the endpoint that used for other drugs of addiction. But the FDA has also added now percent subject with no heavy drinking days. That's a mouthful. Sometimes I abbreviate this PSNHDD.

Basically, it's pretty simple. It's basically just the number of heavy drinking days,

Page 99

no heavy drinking days divided by the total number of subjects during treatment. Again, just to remind you, a heavy drinking day is 4 or more drinks per drinking day for women and 5 or more drinks per drinking day for men. Also to remind you, no heavy drinking days includes abstinence as well as low-risk drinking. And low-risk drinking is those who didn't meet the heavy drinking.

So the question is, what evidence was provided for consideration of this percent subjects with no heavy drinking days as a primary endpoint? This really came from three different types of data sets: clinical trials, treatment settings, as well as epidemiologic studies. And I'm going to talk just briefly about each one of those and just kind of highlight some of the studies that have been done for those.

Before I get into the analysis that we did, primarily Dan and our group did, the combined trial, which most of you are probably familiar with -- it was a 4-month treatment study, 1-year follow-up. It's very large. In fact, it was the

Page 100

largest one we ever conducted. It was 1383.

It's a great data set to work from. In fact, I've given it out now to 36 different research groups around the world, and we still get a lot of publications from that combined data set. At randomization, they were required to have anywhere from 3 to 21 days abstinence, and these were the medications and the behavioral therapies.

The other thing I just want to introduce you to briefly is this Drinker Inventory of Consequences. These are basically the alcohol-related consequences. It's given the acronym DRINC, with a C, and that was developed by Bill Miller a long time ago. And Kathy Carroll looking, "Yeah, I remember that." And it was developed for a project match.

Basically, in our analysis, we used a 37 item or the four subscales: physical, social responsibility, interpersonal, and impulse control. It actually has another subscale, intrapersonal, which makes it 45 items. We thought that wasn't quite as important for what we wanted to look at.

Page 101

Though, to be quite honest, if we did include it, it didn't make any difference in our results.

Let me just go through first the data here. This is -- again, the Y-axis is the drink scores. And of course, the higher the drink score, it means more consequences you had. We took the last 2 months of treatment -- that's actually where we found the biggest effect of naltrexone combined.

You can see the red here are those who had no heavy drinking days, and they had a value round 2.5, whereas those who had any heavy drinking days had an average of 13.7. So you can see they were suffering a lot more consequences. And we thought, well, it would be nice. Let's just follow up on these patients to see if this last. And 2 and a half months later, still a big difference between the two; even at 9 months, and a year later. It was certainly a big difference between the two.

We also thought it would be interesting, well, you know, since no heavy drinking days, consist of abstinence and low-risk drinking, it would be interesting did they drink; did they

Page 102

differ much. So we looked at the DRINK scores here. The light blue is the abstinence, and the darker blue is the low-risk drinkers. Of course, the yellow again is the heavy drinkers.

You can see at the end of treatment that they were very close in terms of their consequences. And again, follow up two and a half months later, it did go up a little bit, but certainly they were pretty close. And it's definitely different from the heavy drinking.

As you go up further, of course they no longer got the treatment. They slight increased, but still they were fairly close together. So generally that's a trend we find. A low risk may be a little bit worse that the abstainers in terms of returning to drinking for long term.

We also just looked at the other one instead of the consequences, what about if they took a drink on a drinking day? How many drinks did they have? Again, at treatment -- of course, the abstinence didn't have any drinks than the treatment. Those at low risk only had 2.4 drinks

Page 103

when they did drink. And those who had any heavy drinking averaged 8.2. And you can see as we follow up later, they weren't too far apart, the abstinent group and the low risk. It's a little bit higher, but certainly different from the heavy. And this sort of continued all the way out to 9 or 12 months. Also, I might mention they did some analysis on Project MATCH, and we really found pretty much the same results in these clinical trials.

Now, let me just briefly just review some of the data we found in the treatment setting. We worked with Connie Weisner, who's here today. This is Connie's data and her group. What they did was they took two data sets and roughly 995 patients. And what we did was after they were initiated in treatment, we looked at them 6 months after the initiated treatment. And then we classified them into three groups: the abstinent group, the low-risk drinkers, and the heavy drinkers. Again, we went back 30 days, so it was at month 6.

Then we wanted to know, okay, what did they

Page 104

look like 6 months later, say 12 months after the initiation of treatment. And we looked at their drinking as well as their consequence, and also looked up to 5 years looking at their treatment utilization and cost. I just want to now give you a summary of it, and I'm sure Connie can go into a lot more detail about those analyses. We actually ended up getting two publications out of it.

This is the bottom line on this. Again, we compared low risk to abstinent group, and then heavy drinking to the abstinent. Again, they were determined 6 months at their initiation of treatment. Then we looked at them at 12 months. We found that going back to drinking was greater in the low risk than the abstinent, but it was much greater in the heavy drinking. In fact, the drinking was closer to the abstinent group, and this really did separate out a lot more in the heavy group.

We looked at consequences, particularly the psychiatric, family, social problems, and they were somewhat similar to the abstinent group, the

Page 105

low risk, where the heavy drinking was higher, significant higher than the abstinent group. We also then looked over a 5-year period treatment utilization, particularly looking at inpatient and emergency room, and we found they were similar, low risk to the abstinent group. But it was significant higher in the heavy drinking. And then finally, we looked at the treatment cost against similar in the low risk but higher in the heavy drinking compared to the abstinent group.

Let me just review one other set of data that was looked at. These are just results of two epi studies. One was by Tom Greenfield. And this is something we did a contract for him to do this. He did it actually on two of his national alcohol surveys. He's actually published the first one on general population, but the FDA was really more interested in these treated or concerned drinkers.

Tom told me he's working on that paper. Sometimes he needs a little pushing. He did say he added another survey to it. But this is what he found so far, was that basically that those who had

Page 106

low volume drinking and did not any heavy drinking days, had a low risk for alcohol dependence and abuse. If they had any heavy drinking day, you really did see that go up.

Finally, we looked at the NESARC data, which is something NIAAA supported. It's a big one, over 40,000. These are just some results that Deborah Dawnson and Bridget Grant and that group came up with. And basically, what they found was subjects with no heavy drinking days carried a much lower risk for alcohol dependence and alcohol use disorder symptoms than those who experienced heavy drinking.

There were actually even some more data that I'm not presenting I think that Celia had in her draft. But really, if you look at the summary of this, if you look at no heavy drinking versus heavy drinking, no heavy drinking decreased the risk for relapse to heavy drinking, dependence, consequences, treatment utilization, cost, compared to those who had heavy drinking. And even looking at abstinence versus low risk, versus heavy

Page 107

drinking, the relapse to drinking dependence for heavy drinking was much greater than the low risk. And it was somewhat greater, though, than the abstinent group.

In terms of consequences, the heavy drinking was different than the low risk and pretty much the same, what we found so far, with the abstinent group, as well as treatment utilization and cost the same. The heavy drinking was greater than the low risk and pretty close to the abstinent.

So if you integrate all this data together, it really had the FDA and their guidance -- basically were saying that patients who never exceeded the heavy drinking limits had minimal alcohol-related consequences and were much likely to relapse at follow-up -- less likely.

Thank you, Dan.

See, this is why it's good to have Dan around. We work so well as a team that when I get off track, they always get me back on again.

A lot of our investigators and other people were saying, okay, you've got this no percent

Page 108

subject to no heavy drinking days. How sensitive is this? We tried to go out and see if we could analyze this. I know we've looked at -- our group's looked at at least a dozen data sets from 12 or so multisite clinical trials for alcohol. We particularly looked at the ones where we actually found an effect.

Yeah, it's hard to believe. We were not positive in our trials. But all we said, if you're positive in all your trials, you'll never believe us. So it's always good to have a couple negatives in there. It kind of gives you credibility.

Anyway, we found that so far from what we've seen, it's not quite as sensitive outcome measures, which isn't surprising. And particularly the number of heavy drinking days, average seems to be a little more sensitive. But it was certainly significant in 5 alcohol clinical trials. And when we tested this in this outcome measure, we found that it was only really significant in two of the trials.

I want to give you some examples here, going

Page 109

back to a COMBINE trial, where we did find it was sensitive, and look at one like the varenicline trial, which was less sensitive. And we're going to compare this with the continuous outcome measures, which I'm going to show you. And then I'm going to compare it with the abstinent outcome measure. So I'm going to do that.

This is the COMBINE study, again, the continuous measures, the really popular one, the average number of percent heavy drinking days, drinks per day, drinks per drinking day, and percent days absent. And they all showed a significant effect. By the way, this was the last two months of treatment, is where we got a bigger effect, though we did find similar results if you go back three months. And if you look at the dichotomous measures, percent subjects no heavy drinking day, we were picking up a significant difference in that. Percent abstinence, we didn't quite pick up a difference in that.

But I want to talk a little bit more about these dichotomous measures and introduce the

Page 110

concept of grace period to you. A grace period is a period of the trial where the outcome is not considered in the analysis because the measured treatment effects is not thought to be representative of the full potential of the drug and the pattern of drinking may still be unstable.

In other words, you may have a 6-month trial. You might fail 3 months. You could have a grace period because things weren't settled down. The drug may not be working yet or the pattern of drinking wasn't stable enough to get that. And that's also in the draft guidance that the FDA put up, that you can have a grace period. But you had to defend what your grace period is going to be.

With that, let me go through and first talk about total abstinence. Here is percent subjects with total abstinence. The blue is the naltrexone, and the yellow is the placebo. This first one here on the left is of the whole trial, the first 4 months of treatment. And you can see, with the effect size around .07, it wasn't significant.

Just looking at the last 3 months of

Page 111

treatment, this separated a little bit more. The effect size went up a little bit. It still wasn't significant. At the last 2 months, maybe a little bit better. And the last month, you hit the jackpot. We actually found a significant difference in that between the naltrexone and placebo.

Now, let's look at the percent subjects with no heavy drinking days. The percentage, again if you take the full stud -- and again, the blue is naltrexone, the yellow placebo -- there was a separation, effect size around .13, but it wasn't significant. If you go back and look at the last 3 months, we were getting significant; .17 started to become significant. The last 2 months was even more, and the last month gave us the biggest effect and had the biggest significance.

Again, just to compare, the red here is the effect size with the total abstinence. You can see that over the duration, the effect size was more with the percent subjects with no heavy drinking days than it was abstinence. So at least in this

Page 112

study, it appeared that the percent subjects with no heavy drinking was more sensitive than the total abstinence.

Now I'd like to talk about our varenicline trial. This was a multisite trial that we did, that we published. It was a 3-month trial. Alcohol, we had about 198 subjects. An interesting randomization, we didn't require any abstinence. I don't think that matters too much because we're getting positive effects and negative with both abstinence or non-abstinence. What matter is we noticed the placebo effect will be smaller, particularly the first couple of months, than the abstinent. Six months is usually the required trial for a pivotal trial. And we're not quite sure how that would play out over six months. We do find the placebo rate does go up each month for those who were drinking up to randomization.

Anyway, the results, what we found was that there was a significant reduction on many of the continuous measures. It was not significant in the dichotomous measures. And just to go in a little

Page 113

more detail, these were the continuous measures. Again, this was taken over 12 weeks. It was actually week 2 through 13. We had a first week titration. We found a significant effect for percent to heavy drinking days, which was, by the way, our primary outcome measure for this one. This was a proof of concept trial; proof of concept trial, you can do anything you want.

Drinks per day, we found an effect, drinks per drinking day. Interesting, percent days abstinent, we did not find an effect. And we looked at the dichotomous measure, and we didn't find an effect for either one of those two.

I'd like now just to discuss these two measures in terms of grace periods. Again, this is the percent subjects who are abstinent. These values are over the whole maintenance period. There was no difference, no effect size, and the values were smaller, too. I think if you remember COMBINE, they were around 17 percent, somewhere like that.

If you look at the last 2 months, the values

Page 114

actually go up, and you start seeing a separation, that it's not significant. And the last month, though, you really did start seeing separation between the varenicline and the placebo. It's close to significance if it wasn't. So if this was carried out a little bit longer, maybe 4 or 5 months, we might have been able to pick up a significant difference here.

Again, just to compare the effect size, the red here is total abstinence, and this is percent subjects with no heavy drinking. You can see that, again, in this study, it appears that the percent subjects no heavy drinking days was a little more sensitive to that.

In summary, as an endpoint, it's probably not as sensitive as some of these continuous outcome measures. It appears to be more sensitive than abstinence, though we have to really check this with more trials, just the ones we've seen so far. It appears you definitely need a grace period if you're going to show significance in these dichotomous measures.

Page 115

Now, when we first had this present subjects with no drinking, people were coming to us and saying, "Well, what are you doing?" I said, "Blame Dan. He's first author on this." But we said okay. What would happen if you had one heavy drinking day, or 2, or 3, or 5, or 10, or whatever? Because people say they relapse. Why don't you allow some heavy drinking? So we said, okay. Let's look and see what happens if you allow heavy drinking, and what will that do to the difference between the two.

To do that, we decided to do an analysis called, a Cumulative Proportion of Responder Analysis. That has been around. It's been used for other medical disorders for pain. In fact, the FDA has actually even used it in their insert package for pain. But it's never been done in alcohol, and I'm not sure if it's even been done in addiction, this type of analysis.

So what is it? Well, it represents a proportion of responders over the entire range of possible cut-off on the graphs. You can take the

Page 116

number of heavy drinking days and create all kinds of different situations, zero heavy drinking, 1, 2, or whatever.

Instead of explaining all this, let me just show you what it looks like in the graph. This is really a simple graph or else I've looked at it so many times, it looks simple. I'm not sure. But I think it is pretty simple.

What it is here, on the X-axis is the number of heavy drinking days allowed. You go from here, zero heavy drinking, or you could allow up to 1, 2, 10, 20. Again, this was I think the last 3 months of treatment here. So you could theoretically have up to 90 days of heavy drinking, though nobody really did that.

What this purple is here are the placebo values, and it tells you the proportion who had no say, no heavy drinking here. And the blue here is the naltrexone values. The red is the effect size between the two values, and that's the effect size here.

So let's start with zero, which is no heavy

Page 117

drinking days, which is the outcome measure. And going back 3 months, the placebo had roughly 30 percent and the naltrexone, about 38. But they had an effect size slightly under .2. So if you add 1 or 2 -- say 2 heavy drinking days, if you allow that, we notice that the effect size goes up a little bit here, over .2. And it seems to peak around somewhere between 10 heavy drinking days, allowing 10 heavy drinking days, or allowing 20 heavy drinking days does seem to be the optimal. And then it seems to go down from there.

So you say, well, why not allow 10 heavy drinking days or whatever. Well, there is a price to pay for that. If you're allowing more heavy drinking days, the consequences are going to increase. And this is an example where we plotted the number of heavy drinking days with the drink score. And you can see, even going up 1 heavy drinking day -- and again, I think these are the past 8 weeks in this one. But anyway, just jumping up to 1 heavy drinking day, you do get an increase in consequences, and it goes up, as it should if it

Page 118

was a good measure. As you drink more, you get more consequences with that.

So you'd sort of have to justify, if we were doing 10 heavy drinking days, what is the clinical benefit of doing that. Plus the fact it may be a little weird if you had that in your package insert, and a physician said, "Well, if I can get you to 10 heavy drinking days, this drug's going to really work." It may not make a lot of sense. But what might make sense, if you do these in different categories and say you can reduce one category to another.

That's sort of a lead-in to my next part, is what new analyses are being conducted to possibly expand the primary endpoints for alcohol clinical trials? We're really thinking about three areas here that we're trying to develop and validate. One is can we validate and show clinical relevance in reduction of continuous drinking outcome? Secondly, can we develop categories of drinking levels and patterns? And third, can we develop some non-drinking outcomes that are sensitive?

Page 119

So with that, we're looking -- by the way, this is something in progress right now, but I'm going to give you an update where we're at on this and what we're thinking. And this is really the data sets we're really interested in. Jurgen Rehm has this chronic disease. He has a lot of data on that where you have amount of drinking and the risk for chronic disease. He has at least 15 or more that he's done on that, and I'm going to describe that in a second.

Certainly, our clinical trials are COMBINE, MATCH, and what we call the NCIG trials. we have a network of sites that we do clinical trials. We finished four already and starting a fifth one. Also, the NESARC -- in fact, there's new NESARC data that just got completed, or study just got completed, as well as perhaps even looking at some other large epidemiological surveys. We're also interested in Connie Weisner. We like her data sets, and perhaps even some other HMO research networks.

Let me just move on and talk about what

Page 120

we're thinking about this continuous drinking outcome. If you're going to validate this, we think there are a lot of things you could validate it; certainly the alcohol-related consequences, treatment utilization, treatment cost. Maybe some others we haven't thought of.

What we have done, we know that a Rehm study has done a lot of analysis on amount of alcohol consumed and various chronic diseases. And these are long-term. These are what Ashley would say is a distal. Rehm has sent us equations so that we could plot these, where we could plot these with the amount of drinking versus the risk of these various diseases. And this is an example, some curves we got for cancer of the upper digestive tract and the lower.

Let me just take this a step further. For example, suppose we take the one for mouth and oral cancer. And this is the amount of alcohol consumed, and this is the relative risk of the disease. Then we took our varenicline study, and what we found was in the treatment, the varenicline

Page 121

group was down to 4.4 drinks a day, and the placebo was 5.4. So it's a difference of 1 drink. And you might say, well, one drink doesn't seem like much. I always like to think, well, maybe if you take it per week, that's 7 drinks a week, or if you take it per year, that's 365 drinks a year. So that sounds a little more impressive than 1.

Anyway, if you take this graph here and put where the placebo would be, at the 5.4 drinks per day, and the varenicline at 4.4, and look at the risk factor, the placebo has a risk factor of 4.4, and the varenicline has it at 3.5. So by reducing one drink, you reduce the risk factor from 4.4 to 3.5. Is that significant? Well, maybe.

Rehm has now, though, integrated all this data and has now put it into a table. This was just recently published. And to be honest, we're still studying this, and we want to study in a lot more detail. But basically, what he has done -- and it's really neat -- he has taken the amount of alcohol consumption per year, and then he has made a table where you get a number of events

Page 122

per 100,000 patients.

For example, he has 15,000 to 18,000 grams of alcohol per year, and he has a risk of development of ischemic heart disease, stroke, traffic injuries, other injuries, cirrhosis, pancreas, and so forth. So he's developed these tables, and he's also developed tables for heavy drinking, the number of heavy drinking days per year.

Now, what's interesting about this is that Lundbeck got approval in Europe for nalmafene, and they did this in three very large clinical trials. And they were using a continuous measure of reduction in drinking and reduction in heavy drinking. But to show that there was a clinical benefit, the EMA actually allowed those tables.

This is an example. The difference between nalmafene and placebo, as an average, was around one drink a day, and that's somewhat of a conservative value. But with that, looking at these tables, they found that it was 692 fewer alcohol trivial diseases or injuries per 100,000

Page 123

alcohol-dependent patients. And he's done this with other different categories. For heavy drinking, he found -- it was pretty conservative, but he found that they had 3 heavy drinking days less a month, the difference between nalmafene and placebo, and then they had 941 fewer diseases and injuries and so forth.

We're still studying this, and Dan's going to talk a little bit more about this in his discussion. But we had to also know that there are weaknesses in getting this data. So we're evaluating this in how they came up with these tables, right now.

The other thing we're interested in is developing risk categories for alcohol intake and can we establish categories describing the risks-benefits at different levels of drinking and maybe patterns of drinking, sort of similar to what's been done in clinical categories developed for blood pressure, cholesterol, for diabetes. This is an example of blood pressure. You could find that. They may even have a stage 3 now. I'm not sure.

Page 124

They keep changing this a little bit. But normal, pre-hypertension, stage 1, stage 2, where the blood pressure goes up.

The question we had asked, well, can we develop these, too? Dan and I sat around one afternoon and made up this table, so it is kind of made up. You have abstinence, low-risk drinking. In fact, you might even have two different levels of low-risk drinking. And then you have that break off with no heavy to heavy drinking, where you have high risk. Perhaps you have different stages: low, moderate, and severe. But the challenge we have to do is to fill this in. These have to be filled in with clinical relevance from different types of data sets, to fill this in to give it some meeting.

I think doing something like this would be extremely valuable, not only to regulatory agencies, pharmaceutical companies, and researchers. I think this would be very useful to patients, be very useful for clinicians, and also third payers. Insurance companies would like to

Page 125

see if you're going from one category to another, what is the advantage of doing that? How is that saving you money?

Pharmaceutical companies, whenever you go to a pharmaceutical company, the first thing they want to know is -- well, first the thing is, can we make money off of it? But then they say, well, what is the clinical benefit of how you do these studies? And they won't even begin to do any drug development until they do that. And I know pharma people, Amy and others, could vouch for that.

Anyway, that's something we're thinking about. I'll just mention, though, the WHO has different criteria that they develop. I had to tell you though, there's probably not a lot of validation for how they came up with these numbers, to be honest. But it's interesting that the EMA has accepted the fact that one of the outcomes could be if you reduce two categories. And most of them come in high or very high. So if they went from very high to medium or low, it would count as a success, or if you come in high, you could be

Page 126

down to low or abstinence.

I'm just about done. But anyway, there is some indication that people could use something like that.

The last thing -- and we'll finish up here -- is just on these non-drinking outcome measures. So far, we haven't found them to be generally as sensitive as drinking measures, or they're certainly not more sensitive for sure. We are interested, though, in trying to validate, through the FDA, two of them. One is the alcohol-related consequence, the other one craving, which we haven't really started yet. We might even have to change the name from craving to urge or something.

But the alcohol-related consequence, Lily started validating the drink. This is something Bill Miller made up, sort of like what Joe Volpicelli did. So they really started to validate this. They interviewed patients, researchers. They did factor analysis with COMBINE, MATCH. The person who's really in charge of this at the time,

Page 127

we're very lucky because she's here today. Amy Duhig was very much involved with that program. Of course, she can answer any questions.

What Amy and the group came up -- by the way, they lost interest, really, so they gave us the data to go to the FDA to see if we could -- and I've submitted something to Ashley, and I know she'll get back to me soon on that. I have a full plate, believe me, but people are asking me about it.

Anyway, they called this IMBIBE, and it's 15 items. We actually did this, IMBIBE, in our varenicline study. Just to show you some analysis -- by the way, we didn't get a difference between placebo and varenicline on the IMBIBE. What we did was take the last month of treatment -- actually, varenicline seemed to work -- got stronger as it got further down the trial. It took at least halfway through before we really began to see some effects with it. But anyway, this is the number of heavy drinking days versus the score. And as predicted, as you had

Page 128

more heavy drinking days, it did go up.

So we thought, it would be fun if we took the average value of the placebo at the end of treatment -- so the average placebo, the last 4 weeks averaged around 12.5 heavy drinking days, and the average value for the varenicline the last 4 weeks was around 8.5. So it did go down a little bit. But I think it's pretty obvious why we didn't get a significant difference between the two. Now, if the drinking had gone down in here, we might have been able to pick up a significant effect.

We also really continue to look at this. We're looking at individual items and how they vary with alcohol consumption. These are the ones that vary the most. I would tell you what they were, except I can't read them from here. But anyway, we're looking at that. We're also doing a lot of exploratory analysis with NESARC and some of the other data sets to see if we can find consequences that are really sensitive to the various drinking outcomes and patterns.

Basically in summary, I think we're making

Page 129

progress. We've still got a long ways to go, in developing and evaluating new sensitive and clinical relevant alcohol outcomes. These are the three categories right now we're looking at. And we're hoping of course for any new approaches over the next decade.

So with that, I think I'd like to introduce my colleague, Dan Falk, who will now tell you -- I don't know if it's the dirty side of this, but will tell you things aren't easy to do, but he's going to tell you the nuts and bolts about how to do these types of analyses.

(Applause.)

Presentation - Dan Falk

DR. FALK: Hi. I'm Dan Falk, NIAAA. So I just thought I'd talk briefly about some key issues just to dovetail on what Raye was talking about.

First, just backing up what makes a good outcome measure, and then we're going to talk about sensitivity of alternative drinking outcomes, as well as the non-drinking outcomes. We'll talk more details -- the devil's in the details -- when it

Page 130

comes to validating these outcomes. There are a lot of data issues. And what are the most important validating outcomes, how do we decide upon them, how do we integrate the results. And then I'll try to jump off for challenges for -- I know what you all care about most are stimulant trials.

This is kind of what we think makes a good outcome measure. First, it's got to be clinically meaningful. So if it's response variable or dichotomous outcome, it should set a threshold by which a clinician might judge a patient to be well or have gotten better. It should be able to be achieved by a sizeable proportion of subjects, and that's not always easy.

In the topiramate trial, we found that no subjects achieved abstinence in the full maintenance period, and only 5 percent in the topiramate group. So it's kind of questionable whether it's a good outcome for all populations or trials.

Number 3, it should be validated or

Page 131

correlated with other informative outcomes, so that's the bulk of the work that Raye has been talking about. You can validate against alcohol-related consequences, physical health markers like blood pressure, global indicators of well being like the SF12. And then finally -- well, then next, it should be sensitive to the effective medication. I mean, that's obvious, but that's critical.

We tend to get really excited, thinking about all these potential outcome measures we can come up with, but are they going to be sensitive? Well, let's see. So they should be at least as sensitive as other outcome measures or at least no less sensitive.

Then finally, and very importantly, it should have the support of the key stakeholders because we've learned in the active group that we're part of, different stakeholders have very different perceptions on what's important.

So first off, the sensitivity of alternate drinking and non-drinking outcomes. A lot of

Page 132

people will come up with these outcome, but I say be careful what you wish for. You might get it. Some of these are not as sensitive. These alternative drinking outcomes can be less sensitive.

This was what Raye was talking about. This was the EMA, the FDA equivalent in Europe, came up with this reduction in drinking by two or more levels, from baseline to treatment. So you can go from very high to medium low. Abstinence, that's a two-shift or from high. These are most of where our people are coming in our clinical trials. You can go to low or abstinent.

But the real question is, is it sensitive? This was in COMBINE. You've seen all this before, and I just added this at the bottom. It's significant, but the treatment effects are a little less than percent subjects, no heavy drinking days and less than the continuous outcomes. Now, we still have to test it in other trials, but it may diminish some enthusiasm. I'm not sure.

Also, there are other dichotomous endpoints

Page 133

that can be proposed, and the EMA proposed several more. They proposed a proportion of subjects with a reduction in alcohol of actually 50, 70, 90 percent or 100 percent. A hundred percent is total abstinence.

So we're borrowing from our friend the Cumulative Proportion Responder Curve here. This is from a topiramate trial. This was in one of our papers. So Raye introduced how to read this. But basically, on the right-hand side where it says a hundred percent reduction, that's total abstinence. You get a treatment effect of about .43, let's say.

The other arrows, that's the 90 percent right there, that's 75, and that's 50. What you see, there's some fluctuation there. Ninety percent of the treatment effects drops; kind of picks up a little bit with 75 percent. But then by 50 percent reduction, it goes down to about .3. So you do go from .43 to about .3, so you do lose some treatment effect there by going to this more lenient -- let's call it more lenient responder definition.

Page 134

This is just one more of these. The X-axis is drinks per day. On the right side where it says zero drinks per day, that's total abstinence. This is in the topiramate trial. Again, with total abstinence, you get .43. What happens if you -- let's say you wanted to create a dichotomous outcome at 5 drinks per day. These people that enter these trials have maybe 8 to 9 drinks per day when they come in. So you might think, well, 5 might be pretty good. They're cutting back to about 5 drinks per day. Let's call that a good outcome if they have 5 or less.

But if you did that, you would see that the treatment effect would kind of plummet there from .43 to about .17 or something. So you kind of have to be careful what you wish for because you could get it if you choose these more lenient outcomes.

Let's talk about non-drinking outcomes. These were all drinking outcomes. What about the non-drinking outcomes? FDA has expressed in their guidance that drinking is really a surrogate for the non-drinking outcomes, and we've heard that a little bit this morning. The aim of treatment is often expressed as an effort to modify drinking behavior, but the actual desired effect is improvement in physical and social consequences, in the non-drinking outcomes.

Page 135

So we would love to measure the non-drinking outcomes in our trials and have them be significant. That would be wonderful. We wouldn't even have to bother probably validating the drinking ones, right? But the problem is that they may have limited sensitivity or at least variable, and that's what Raye I think was alluding to there.

LoCastro et al. in 2009 looked at COMBINE, and he basically found -- these are all secondary outcomes, all non-drinking outcomes. There's a wide variety here. And most of them were not significant, with exception of a couple here, SF12 health score and maybe a WHO, environmental. But actually, they concluded -- they did so many tests that after they controlled for multiplicity, they found that none of these were significant. We did a couple more analyses on the drink score; again,

Page 136

not significant, and blood pressure wasn't significant either.

So that's the challenges, that we can't find significant things. In other trials, I think Bankole Johnson in the topiramate trial found that craving and drink score, the consequences were significant. This is just one trial, a large one. So that's our challenge here, is that we can't really find significance in these. That's why we have to validate the drinking outcomes. And that brings us to the issues of the devil's in the details when you're validating drinking outcomes.

The validation of our drinking outcomes is really only as good as the data against which we validated. And each of these data sets here at the top have strengths and weaknesses in terms of -- so here's our drinking outcome. Well, the Rehm data tends to just look at total alcohol consumption, but we know there's a lot of other drinking outcome measures we might want to validate, like the number of heavy drinking days, the percent drinking days. That's not here in these data. But we do get it in

Page 137

our clinical trial data, where we have very -- we use the timeline follow-back. We can calculate any kind of outcome we want from the daily records of drinking.

NESARC has pretty good, and the Kaiser has pretty good drinking as well. In terms of the consequences, there's chronic and acute consequences. The Rehm data is very good with chronic. In our trials, we don't have any real chronic data that we collect. Even if we did, it would probably be too short of a duration to collect it. But NESARC has a nice -- they measure chronic and acute pretty well, and so does Kaiser. We do measure in our trials acute consequences very well with the drink instrument.

In terms of study design, ideally, you'd want a longitudinal study design. When you look at the drinking outcomes at wave 1 or time period 1, you look to see how it predicts consequences later on. And that's the strongest test, and that's what Ashley was talking about. That was one of her spokes, was a longitudinal validation, and that

Page 138

seemed to come after doing maybe a cross-sectional one.

The cross-sectional -- the Rehm data -- we're trying to understand how they churn the numbers out in the Rehm data, but it seems to be a little more cross-sectional, but that needs to be confirmed. But at least with these other data sets, it's more longitudinal.

This is blood pressure. The question is which outcomes do we validate and against what consequences? This is blood pressure. They've got it pretty easy in some way. They have two outcome measures, systolic/diastolic, and they're going to kind of crunch these numbers, maybe come up with the cut-offs, 140/90. And those are going to be validated against these consequences, which are kind of limited.

This one comes from their most recent 2014 and the JNC, committee, which develops the blood pressure guidelines. So they said that this is what they look for as outcomes to validate blood pressure. It's basically just cardiovascular

Page 139

health outcomes and kidney, renal outcomes.

But what's it like for drinking? We have at least 7 to 10 different types of drinking outcomes we might want to -- these are continuous measures. And then we're going to slice and dice them an infinite amount of ways just with the number of heavy drinking days. You could have no heavy drinking days, 1, 2, up to 3, 5, 10, whatever you want. But then we're going to have to validate these against -- there could be like 65 consequences that -- because alcohol's a dirty drug. It affects all these physical, these health related aspects, as well as the drink has 45 items in these 5 classes. We've got interpersonal, economic, employment, all these other consequences.

So it's really a challenge. This is the challenging question, as I see it, is how do we integrate the data? How are we going to validate a given clinical trial result on a continuous outcome? First off. And then how are we going to create and validate a responder definition given the myriad consequences by which to validate them?

Page 140

In the very limited -- it's a simple example. I only picked two outcome measures, two health consequences. They're both cancers. One is occurring a little higher up in your mouth. The other one's occurring a little lower in your esophagus. And you have the alcohol consumption, the outcome measure you'd like to validate, and the relative risk.

What you notice that's different about these is that the slope of the risk curve is different. This is steeper here than this one, which means that, basically, you start seeing risks occurring with lower levels of drinking.

Let's take the blue. We're going to validate a clinical trial result. That's what Raye was talking about when you're comparing placebo and medication on the actual outcome that you achieved in your clinical trial. We said 4 here, 4 drinks per day, in placebo; a 1 drink difference corresponds to 1 relative risk unit different, let's say, on this outcome. But on another outcome, really there may be very little relative

Page 141

risk difference in the clinical trial result we see.

So the question is, well, which outcome do we use? Now, this is only 2 outcomes. There are going to be 65 of these curves even in ideal state that we would eventually develop. How do we know which ones to look that? Do we pick the one with the steepest curve if we want to be really conservative with what we're doing? We want to minimize any risks, so maybe we take the one with the steepest curve.

Now, if we want to create and validate a responder definition, we could decide a priori, we will not accept any amount of alcohol in our responder definition that increases from zero to a relative risk of 2. So we're doing to cap at a priori at a relative risk of 2. If this responder definition is going to increase our relative risk more than 2, forget it.

So let's say that's how we're going to do it a priori. Now, do we choose 2? Do we choose 1.5? Do we choose 3? We'd have to have a debate about

Page 142

that. But if you did, you'd see that the relative risk of 2 does translate to different amounts of alcohol consumption that you'd make your cut-offs.

I don't think we have an answer to this, but I do know in the blood pressure -- I was trying to read how they came to consensus. They had big meetings. I think they invited 50 -- I don't know. I didn't tally up all the numbers of experts. It looked like 20 to 50, and they had special -- I met the people, NHLBI, that set up these meetings. They actually hired a special consultant who was expert in achieving consensus, get these people together, herd the cats, and try to come up with -- and they reviewed -- each expert reviewed the meta-analyses. But they didn't have the challenge that we had. I don't think they had as many outcome measures, and they did not have as many consequences to look at.

Now this brings us to the challenges for you all for the stimulant trials. I don't know much about stimulant trials, but this is what our understanding is. I don't think you have timeline

Page 143

follow-back like we do, where you measure daily quantity. Do you have daily quantity of -- no. So that's correct. There's not daily quantity of I guess cocaine amount consumed.

Okay. Two minutes. That's fine. So maybe you only have frequency like percentage of days that are abstinent or maybe just total abstinence, right? But if you only have frequency, you need to evaluate the clinical relevance, develop sensitive non-substance, intake endpoints like the health -- all these different kinds of consequences. But definitely before going through the process of picking the ones you like, make sure that they're sensitive, and try to testing as many clinical trials as you can.

That's our very low level of advice. We're not experts, but that's kind of where we're at. So that's it.

(Applause.)

Q&A - Group Discussion

DR. STRAIN: If it's all right, I'd suggest that we maybe start lunch a little later. We've

Page 144

got plenty of time for lunch, and you can all eat fast, because I think these were really a great pair of studies.

Raye, do you want to come up as well? I think it would be useful to take a couple minutes to see if there are any questions people have because this is such a -- as sort of a model for stimulants, this seems like a critical opportunity to look at it.

George?

DR. WOODY: George Woody. One of the slides said GGT, but I had the impression that liver enzymes are a pretty sensitive measure of change, but I didn't see that coming up too much in any of these studies.

DR. LITTEN: Yes. George, we generally do use biomarkers in our trial. The liver enzymes just aren't very sensitive. We're more, though, excited about these new alcohol metabolites: ethanol, glucuronide, and phosphatidyl ethanol, particularly because they're more direct metabolites.

Page 145

But when you use biomarkers, there are limitations to them. You have to know what the limitations are. First of all, biomarkers aren't going to tell you so much how much. It's always yes/no. And you have to know exactly how much you think they're measuring. For example, ethanol glucuronide will measure up to one drink, but it has a very small window of assessment within maybe one, three or four days, depending how much you've drank. Phosphatidyl is a little bit different. That will stick around maybe two or three weeks, maybe four weeks, but you probably had to drink more to get that elevated. So you could sort of use those in combination.

Also, for example, too, it depends on not just how much you were drinking a day. Say you have phosphatidyl, and you need 3 or 4 of 5 drinks to elevate it. It depends on how quickly you did those drinks because if you did it over, say, 12 hours -- in fact, if you had maybe a drink an hour, it may not be elevated. Remember, it's measuring the amount that's in your blood. But if you drank

Page 146

it over, say, a two-hour period, then it would be elevated more.

So you really have to take into account all this. What I think is most exciting for us, and the one we're really trying to get is these alcohol sensors. SCRAM I think has done a great job with that. It's used in the criminal justice system quite a bit. It's around your ankle. But it measures drinking in real time. It's not quite as quantitative as we would like for it to be in terms of blood alcohol levels. But we put out special announcements to improve the technology of that, to make it cheaper, maybe easier to where more quantitative.

Of course, if we start using this, too, we probably have to change our outcome measures, and change it instead of drinks is how much blood alcohol you have because it's really a measure of blood alcohol, not the number of drinks. And as you know, the number of drinks and the amount of blood alcohol you get varies quite a bit, depending on the size of the person, whether they have any

Page 147

food in their stomach, of the enzymes vary that break it down among individuals.

So these are all things worth thinking about in terms of biomarkers. We do like to use biomarkers -- and I think the FDA likes for us to include them in our trial -- more as validation. If we report and self-report that you have a reduction in drinking or whatever, it's nice to have that validated by biomarkers. But you have to realize the limitations of those.

DR. STRAIN: Other questions? Thanks. Ken?

DR. SILVERMAN: This is Ken Silverman. Two terrific presentations. Thanks very much for those. I shouldn't be hung up on this, but I'm hung up on the FDA issues that Eric elaborated on, that you want to show either that the medication improves how long people live or whether how they feel and function in daily life.

You guys described validation of the drinking measures, which seems pretty good to me, but you're focused on drinking as your outcome measure, and you pick measures that are validated,

Page 148

like the drink responses, which seems pretty sensible.

But I'm wondering, is the FDA -- is the focus on how a patient feels and functions in daily life, would that suggest that the drink should be that the validation measures should be the primary outcome measures? Have you given that some thought?

DR. FALK: Yes, exactly. That's what I was trying to show with that one slide with COMBINE, where the drink, the consequences wasn't statistically significant in drink. The effect size was .11, whereas the drinking outcome was .21. That's the problem. And the question is, why isn't the consequences significant in our clinical trials? Maybe it's too short of a period that we're assessing the outcomes. Maybe it takes longer before you finally move the dial in terms of -- like some of the consequences are if you got a DUI, or you got divorced or whatever. It could a while, a lot of reduction before you see improvements in some of these kinds of

Page 149

consequences.
DR. WOODY: Plus different people might show different kinds of effects, different kinds of consequences.
DR. STRAIN: Sorry?
DR. WOODY: Different people may have different kinds of consequences, whereas everyone might decrease drinking -- or more people may. A large percentage, many people decrease drinking as they do in our trials; they decrease cocaine use. But the effect that that has on their lives varies across the people a lot, so it's harder to --
DR. FALK: There's a lot more variability.
DR. LITTEN: And there could be a lag behind this, too. The varenicline was 3 months. If we'd done it 6 months, we'd might have been able to pick up differences in consequences. And we have some data that suggests that you do start seeing consequences and follow more -- after the drinking has occurred.
FEMALE SPEAKER: I think the other thing -- real quick to follow up on that. I was

Page 150

part of the initial -- we took the drink at Lily. And the idea was for us to have an endpoint that would actually resonate with maybe clinicians, but more towards the patient, like what benefits do I actually receive if I reduce my drinking.
So we went through the FDA draft guidance that had just come out I think it was in 2006, or '07, or something like that. So we went back to patients, and we took all the drink items, and we asked them about the relevance, the importance, how often they'd actually experienced those symptoms. And we had to cut out those things that were not -- that were going to take a super long time or the really infrequent items, or things that just really didn't matter.
So we came down with -- it was not perfect, but came down with a shorter list, with the idea of what is relevant and important and what could potentially change. I don't think these trials have never been powered based on the drink either. I mean, you're powering based on your primary, and then hoping that these tertiary or secondary

Page 151

outcomes then move.
Then also, the population that was -- we did the qualitative interviews with about 60 patients, and they were all outpatients, and they were -- with the idea that this is an outpatient drug, with people with not severe comorbidities like would be included in your phase 3 studies, but with the idea that at the end of the day, could you demonstrate value to a patient who said these things were important.
FEMALE SPEAKER: That's the IMBIBE?
FEMALE SPEAKER: That's the IMBIBE, yes. So there are some gaps in there. And if it's just based on the qualitative phase only, I'm not sure that -- then Lily kind of abandoned it when their drug -- so it was turfed over to you guys. But there are so many issues associated with that.
DR. FALK: But even with IMBIBE when we tested in varenicline, where we got nice treatment effects, .3's on most of the drinking outcomes, IMBIBE still wasn't -- it didn't improve our -- our sensitivity wasn't significant. So even with some

Page 152

of these shorter duration consequences -- so it's not clear. Maybe there's more variability. We'd have to check the standard errors and just see if there's more variability around consequence type measures than drinking type measures.
DR. STRAIN: Connie, real quick, and then I want to move on.
DR. WEISNER: I was just going to say, I know it's a power issue, and we've talked about this before, too. But I think we are often selling our treatments short because the issue is that we are finding you see the consequences improve that are related to why the person came to treatment. So unless you're measuring employment outcomes for someone with an employer mandate or legal outcomes for that, you're not seeing this. But again, that's a real sample size issue.
DR. FALK: That's a good point. That's an excellent point.
DR. STRAIN: So thank you. Yes, I know, that granular level to it.
So I want to thank these two guys again.

Page 153

Thank you, NIAAA.
(Applause.)
DR. STRAIN: We're going to move on, but I hope we can continue this discussion over lunch.
Kathy Carroll, you're up.
Presentation - Kathleen Carroll
DR. CARROLL: I'm delighted to be here today. I want to thank ACTTION. And I really want to thank NIDA and Medications Development, Dave McCann and Ivan and Phil for spurring us to do this work, where we've been slavishly following Dr. Falk's work; a big fan! Big fan. This began in about 2011 with a supplement. This is sort of a two-parter.
When we began trying to figure out what's an appropriate endpoint in stimulant trials, we came up with a list of what we wanted. And what I'm going to do today is talk a little bit about the conundrums in selecting stimulant trials or 50 different ways to calculate urines. You think you're doing it all the same way, and you're really not. I've become an aficionado of reading things

Page 154

and reviewing and really looking at those numbers because it's tricky.
Why do we need a sound and valid indicator? I'm preaching to the choir. There's the choir. But we really, really need this in stimulant abuse, and especially as we're moving into primary care medical settings because we can't yet talk about what we do in a way that's convincing to clinicians and payers and so forth. So we absolutely need these things for meta-analyses.
We really do need to set and monitor performance standards. I think something that gets totally in the way of being able to move some of these new treatments into the field is that the field itself, the clinical field of stimulant treatment, doesn't have outcomes. All they've got is retention, and all they've got is group and individual treatment.
So when we come in and say our little CBT for CBT can give you a 30 or 40 percent difference in three or more weeks of continuous abstinence, nobody cares because they don't have a clue about

Page 155

their rate of three or more weeks of continuance abstinence. They don't even collect urines. So we should think about that, too, a clinical field that does not collect urines and doesn't know.
It's complicated, so we really, really do need this. And it's so fascinating. I was fascinated when Raye was talking about the percent of completely abstinent people in alcohol trials. I mean, we should know that. We should know that it's 7 to 9 percent, and we can do better than that.
So anyway, this is something that I think is really important. We need to be able to compare -- let's all take comparisons across a common standard and pull the clinical and research fields together more.
What I'm going to do is talk about what we've been doing in the last three or four years in terms of working through desirable characteristics of indicators. I'm going to talk about some of the strengths and weaknesses of the common approaches for approaching endpoint analyses -- it's really

Page 156

ugly, so the guy's coming back from looking at the truth -- and then talk about our project where we have taken four of our more recent randomized clinical trials of cocaine use disorders and pooled the data, tried to harmonize it, and then go through what it said. So I'll just be talking about how we approached it, and then later this afternoon, Brian Kiluk will be talking about what we actually found.
When we started this, we were also thinking about this approach to what's clinical significance, what's really meaningful. And if you look in the general psychological literature, it's always things that translate to complete abstinence.
So if you look at change of two standard deviations anti-depressant form or something like that, or moving into normative functioning, that's really complicated with cocaine users because, in general, they come in with all kinds of variance in their frequency of quantity of use. So anything that is a change in two standard deviations of when

Page 157

they came in is almost always abstinence. It's a huge standard deviation.

Reliable change indices, just looking at reductions in frequencies doesn't work very much because, again, we're not quite sure what's a safe level or what's an appropriate level of cocaine use. So this notion of return to healthy functioning that Dan and Raye are talking about I think is fascinating, and I think we're just now getting to be able to have the data that can begin to point us in that direction. And I hope that's one of the goals of this meeting, too, what data can begin to inform finding that endpoint for us.

This is what we're looking for in an indicator, and there's a lot of overlap with what Dan talked about. But we want also that's fairly easy to calculate and interpret. I can tell you that no clinician is really impressed with hierarchical linear modelings and the differences in the slopes. It may be great. It may use all the data, but it's not real convincing.

So we want it to be psychometrically sound,

Page 158

reliable and replicable across trials. We were joking about this last night, but low susceptibility to missing data. You can set it up perfectly, but almost all of our really great measures are almost completely undone by missing data, and we do tend to have a lot of it in our trials.

We want something that's verifiable. It's really nice if there's a biological indicator. And we kind of want it to be independent from baseline measures. If severity predicts everything -- and we heard about that a little bit from Kyle, where they come in predicts where they go. We're not really adding much if it doesn't increase our predictive power. It should be clearly sensitive to treatment effects when we've got treatments.

They all cost something. Ideally, it shouldn't cost a lot because we have to remember we live in a field that doesn't collect urines. They don't collect urines. We have to predict long-term cocaine outcomes. That's obviously useful. It should relate to indicators of good, long-term

Page 159

functioning however we define that in terms of work and so forth.

It should be acceptable to the field, and it should be easily interpreted and seen as useful by clinicians, policymakers, and payers. So this is kind of the grail quest for us. And it's going to be hard, I think, to come up with something like this, but it's lovely to begin.

We've been talking about this in different kinds of ways, but when you talk to people in general health care, they say, "Your outcome should be easy." Success in treating substance users should be some sort of durable period of abstinence. They should be working -- that would be nice -- or productive in some way, taking care of the kids. They shouldn't be a burden on the criminal justice system and involved in criminal activity.

It's always complicated when you look at healthcare utilization because a lot of times when they come in to treatment, they're beginning to get involved in preventative care and ideally using

Page 160

hospitalization and emergency rooms less, but not always. I think the only treatment I know -- I was thinking this -- is Ken Silverman's. You get your guys working, and they're not involved in -- you've got it. You probably have the measure of what changes people with your workshop, your contingent workshop.

But we also decided to be brave with this and come up with this as a straw man indicator of outcome in substance users. And we found across our -- it's around 450 patients in this combined data set, that this characterizes about 11 percent of our population at the end of treatment, which is really close to your percent of completely abstinent, and 20 percent at the end of one-year follow-up. I hope that's not lying, but that's where it is.

I think it's useful -- it's sobering to have this kind of straw man indicator. We often find we can get cocaine use to change because that's actually what we're targeting. Our treatments don't necessarily target employment and criminal

Page 161

activity. I think the reasons that some of those measures are insensitive are complicated, as we were just talking about.

One of the things that Brian will talk about just a little bit, too, is why not choose complete abstinence from the beginning to the end of trial for our folks, and this was interesting. It's a relatively insensitive measure. It is difficult for most people. It's about 14 percent in our sample of 434.

One of the things that we found was that the people who were completely abstinent from the beginning of treatment for the full 12 or 8 weeks were not those who actually had the best cocaine use outcomes as they went through a one-year follow-up. There's something odd about those people. And perhaps if you've got a chronically relapsing disorder maybe using a little bit once or twice isn't the worst thing in the world and learning from it.

Again, I've always wondered if you're starting with people who are completely abstinent

Page 162

and it's not a relapse prevention trial, what are treating? They're already abstinent, so that's the reason it may not be terribly sensitive. I like my abstinence slide. It's hard. It's really hard.

So we went through, and we wrote this up in a paper, but when we were approaching how to quantify 15 or 16 different outcomes that we came up with for this project, we just thought about pros and cons for each of them. What's often used, again, in the clinical literature is retention because it's easy. It's really easy to calculate. They're there or they're not, and you can get that on everybody. They're there or they're not.

It certainly is an indicator of treatment acceptability. It can be a really nice indicator of differential attrition and data availability across conditions, too, so there are reasons to collect it. It certainly may be meaningful in some context than others. In our research trials, we find it doesn't predict a whole lot because there's a lot going on in the research context that kind of keeps people in.

Page 163

We pay them for their time doing assessments and so forth, and I think we make research contexts more supportive and willing to be there for the patient than some clinical contexts are, sort of the date, notes, data, and a lot of other data suggests that in the clinical context, sticking around is just about everything. And patients leave treatment for different reasons. Some get bored. Some don't like our treatment. Some of them are around for eight weeks, aren't using, and feel like they're cured; they're done.

There are philosophical things. If you're only looking at retention, that's problematic because is retention with an ineffective treatment really all that meaningful? And again, we didn't find that it was related to long-term cocaine outcomes in our particular sample as well. So, oh well.

Percent negative urines is the big one in our field. We like it. It's widely used. It's accepted. Most people believe in it because it is less susceptible to the demand characteristics and

Page 164

misrepresentation of our patients. With quantitative urines, you're certainly able to detect new episodes. And it can be wonderfully accurate if you set it up right with the right timing and there is no missing data. And then when there are missing data, it's really where all hell breaks loose in the field.

So timing really is critical. Is it twice a week? Is it three times a week? What do you do about overlaps? And we're going to let Kenzie talk about that tomorrow.

This has all been said before, but just to refresh us, for stimulants, it tends to get only recent use only, so 3 to 5 days back. If you're doing quantitative urines, especially 3 times a week, it cost a lot. These kinds of measures are very sensitive to missing data, especially with differential attrition and assuming -- and it's all depending on the assumptions that you make when the data are missing, and I'll get to that in a minute.

It's also complicated because you've got urines, and we tend to focus on just the cocaine

Page 165

dependence, but a lot of times, I think we're a little less sensitive to as the stimulants go away, what goes up? Does drinking go up? Does marijuana use go up? Is that really a good outcome if you get them to stop using cocaine and they're using tons of marijuana, and alcohol, and other kinds of things.

You can't backfill the information that you came with the timeline follow-back. And again, there were all kinds of problems with the assumption that missing is positive. Well, okay. I'll get to that in a minute.

Just to show how -- there are tons of examples of this. Let's say you've got a 12-week trial. As often happens, somebody's around for two sessions. They give up one negative urine, and then they drop out. So there's at least 50 ways you could calculate urines for these guys.

So if you're based on the number submitted, they're all clean. They were abstinent. That's a winner. Hooray! If you based it on possible, it's 50 percent. If you based it on the expected with a

Page 166

one time a week collection schedule, it's 8 percent. If it was 3 times a week, it's 3 percent.

So you have to be really careful. And I guarantee you that when we start harmonizing our data sets, this is going to be a killer because everybody does it differently. And we found that even our data managers, who we all thought were doing it the same way, do it slightly differently because those assumptions are really important.

Longest consecutive weeks, 3 or more consecutive urines, 3 or more weeks. This is a really, really nice one, too, because I think when you get it, it provides strong evidence of meaningful abstinence. It has all the other advantages of looking at urines. Again, timing on this one is so critical. And this one's great, too -- but it's really, really susceptible to the missing data. If you're missing one in between, you've got this urine, this urine, a missed urine, and then another clean one? Is that three in a row that's clean or what happened when -- when they're

Page 167

not around, they're usually using something. There are reasons people don't submit urine.

It sounds great, hard to do. Percent days abstinent through self-report. We do timeline follow-back, and it's widely used. And it's wonderful for approaching these kinds of things because essentially what we have is frequency, not quantity, of cocaine use for the 3 months before somebody comes into our trial, the 8 to 12 weeks they were in the trial, and then we have another -- we have a whole year of data on people.

You can cut that data up any way you want. It's really highly flexible for these kinds of analyses. And you can also do these true intention-to-treat analyses, which we've done with cocaine users. So we tend to -- we get about 80 percent of everyone in the trial and actually do go and find the people who dropped out, which is really nice. And it turns out we can't make that assumption that at the point of dropout, everything's dirty. There are these different trajectories after people drop out. Some people

Page 168

drop out because they're doing well. So you really can't make those kinds of assumptions.

So our general approach has been to look at our discrepancy rate of percent days abstinent and self-report. Across all of our trials, it's about 8 to 12 percent. So if we collect 3,000 urines in the course of a clinical trial and count the number where the urine was positive in cases where the patient did not report cocaine use 3 to 5 days back, it's about 8 to 12 percent. So we've got to have a 10 percent error rate.

Sometimes those errors are -- sometimes we have -- we have liars, really consistent liars and really bad liars.

(Laughter.)

DR. CARROLL: Really, if you look at these things -- we're obsessive about this stuff. But it's interesting. What's also interesting is that rate is somewhat unusual. If you talk to other investigators in the field, their rate of discrepancy of 50 percent. It's really a complicated thing.

Page 169

What are we doing that's differently? We have great research assistance. The thing that we're doing, our rate of discrepancy has gone down since we've been using immediate test cups rather than waiting a week to get the urine results, so the patient can be encouraged to self-correct a little bit. That's another complicated thing.

But percent days abstinent have all kinds of problems as well with higher differential dropout. The denominator gets really complicated as a percent days of the 12 weeks of the time they were in treatment. It's all about the denominators in these kinds of analyses. And it's not always easy to correct the urine data if the discrepancies are really high. So you can certainly take was the patient abstinent for three or more continuous weeks and self-report. And if there's a dirty urine in there, they're a no. But you can't necessarily correct percent heavy days or percent days of abstinence because you could get one or two dirty urines. Is that 5 days? Is that 10? It's really complicated.

Page 170

Maximum days of abstinence overall are in the final 3-4 weeks. We're working at that. That's been nice for us because it has been linked to longer term cocaine use. It's potentially verifiable if the urines are collected at adequate interviews. You can play with this idea of grace period. It's not clear to us yet, and we hope to kind of get back to that.

So we seem to get a signal around this 3 or more weeks of continuance abstinence that Frank Gawin made up umpteen years ago. And it's not clear if it matters to the end of treatment or at some point during treatment. But again, it's really complex when you've got missing data and you've got discrepant urine, which we always do. Are we counting? When you talk about end of treatment, is it end of the patient's time in treatment where you have more data or the end of the actual trial where you have less data? It's complicated.

Reduction in use, frequency and quantity. This is something that I think we have to dive in a bunch of different ways. It is a nice alternative to abstinence. It may be a more achievable target for people. It's also nice because it's really compatible with those lovely random aggression models over time. It may be sensitive to treatments that take a little more time for the effects to emerge. You can get your grace period, and you can also easily dichotomize that kind of thing, too. So you can lay a percent reduction use or the percent of people who get to a 75 percent or a 50 percent reduction in use.

Page 171

As we start playing with it, it's really, really complex because we can't get really good quantitative measures, especially with that baseline period. We don't know what people were doing, in general, the three months before they came to us, and they don't remember real well either. And then, when are you looking for that reduction? Again, in the last weeks, over the entire course? So we could easily get up to 50 or 60 different ways to cut the data here, and we certainly don't want to do that.

Page 172

With stimulant users, some of the things we tripped over, just defining what we want in reduction is really complicated because the patterns of use in stimulant users vary so widely. We've go these huge bingers who can go forever without using, and you get some low-level users, maybe the people who are self-medicating and attention problem, who use a little bit every day.

It's really, really hard -- this came up in Dan's question. Getting these reliable estimations of quantity is just really, really complex. We don't have a standard unit of cocaine. It's an illicit substance. Dealers don't exactly hand them out in standard units. There's all kind of adulterants. The potency varies. The language is really, really complicated, and it can't convert to a dollar value because there's sex for drugs. People share drugs. It's illicit. It's tough. Again, that's where quantitative urines really help, but that's complicated, too.

One of the things that we really wanted to do as we were jumping into the status, that was

Page 173

come up with a nice dichotomous indicator. We know that they're not -- the continuous measures are more powerful for all kinds of reasons, but we guessed at what some reasonable candidates would be, based on the lurcher in cocaine use, alcohol use, smoking and so forth. So we looked at things like complete abstinence, 3 or more weeks of abstinence, end of treatment abstinence, reduction of X percent. And then we had that little straw man of good enough functioning. Are they abstinent at the end of treatment? Are they working? Are they not in jail, kind of thing.

So this is the beginning of the paper that Dr. Strain's journal was kind and generous enough to take because it was so long. Essentially, what we did was take 15 candidate indicators, the most common continuous measures that are used in the field, the dichotomous measures that I just went through, and really talked about them in terms of characteristics that were sort of inherent in the outcome measurement itself, ease of computation, cost, acceptability, and so forth.

Page 174

But some of these also -- some of the things that we're really interested in are questions that can be addressed with data, good empirical questions, so relationship to one year cocaine use, sensitivity to baseline variables and so forth. Brian will be talking about that a little bit. We just, I think, wanted to lay out the conundrums and the complexity of doing this kind of work. But it's been fascinating, though. And we are the Knights of the Holy Grail.

So far, doing this work has also been really, really rewarding and interesting, too. So we've produced a fair amount of data so far and looking forward to do more. That's it.

(Applause.)

DR. STRAIN: Thanks, Kathleen. And thanks to you also -- [inaudible - off mic.] I thought that was Monty Python, actually.

DR. CARROLL: It was Knights of the Holy Grail.

DR. STRAIN: Is it? Okay, good.

I would like to, for the sake of timing, bring up Celia for the discussion, and then we'll take a few minutes for some questions, and then break for lunch. But this has been really helpful, and I appreciate the plug for my journal as well, which is always appreciated.

Page 175

Discussant - Celia Winchell

DR. WINCHELL: Hello. I'm Celia Winchell. I'm the medical team leader for addiction products at the FDA. I'm going to get us rapidly back on track because I don't have any prepared remarks. So I'll just take long enough to say that I greatly appreciate Dr. Carroll and her group and the work that they've been doing, not just the work that they're doing to explore the data, but what a wonderful job they've done in articulating the problem, both here today and in the published papers.

If you have not read them, I very much commend them and recommend that you take a look because one of the major lessons I've learned the hard way, through a few years -- it will be 20 this summer -- of being involved in the review of

Page 176

protocols for these types of treatments, is that there are a lot of great ideas that turn out to be essentially impossible to operationalize. It's a nice idea to say a 50 percent reduction from baseline, but it's very hard to do.

So when we think about what is the best endpoint, we think about a lot of things, what would be meaningful to people, what seems like a good idea, and what can we do. Sensitivity to missing data is very important. And the cost of obtaining the data is also important, if we're talking about development of drugs and what industry would be willing to do.

So if we pick an endpoint that is maybe less sensitive, maybe you lose a little power with a dichotomous endpoint, but you also inure yourself to the difficulties in missing data, or you create a situation where you can adjudicate all the patients, or it's easier to collect the necessary information, and you save some costs that way, and that trade-off might be helpful.

So as we think about what patterns of drug

Page 177

use we can measure proximally that would be valid surrogates for clinical benefit more distally, if we assume we can't do a study long enough to actually observe that clinical benefit, we can also think about those challenges in the clinical trial, the cost, and the difficulty of getting complete ascertainment, and fold those into our discussions. I'll stop there.

Q&A - Group Discussion

DR. STRAIN: So this talk and discussion is open for questions or comments. I'm going to take the prerogative to begin by just making one point. Well maybe I'll make more than one point. I'm really taken with the idea that the alcohol field is sort of two or three steps ahead. So learning from what's worked and what hasn't worked is valuable from them.

In point of fact, Kathy when you made your point about there's a variable use patterns in stimulant users, I found myself thinking, well, oh, gee, that's a problem. And then I thought, wait a second. Alcoholics have variable patterns of use

Page 178

as well. So I think that actually, there yet again, is another parallel there that may be useful to capitalize on, and I wondered if you had any thoughts on that page or the more general point of what can be learned from the alcohol field as you think about the stimulants.

DR. CARROLL: Yes. I think the most important thing is following your lead, which is being guided by the actual data, that you've had COMBINE. The expert panel approach has been useful. We always come up with these great ideas, but it's going to be so important to just imply empiricism to this. And I think that's the number one thing. And we do have enough trials where we can actually do this, and we can turn clinical wisdom on its head a little bit. It turns out complete abstinence may not be all its cracked up to for the cocaine literature, and I think it looked a little less sensitive for the alcohol literature, too.

So can we find our equivalent of no -- I think the approach will be the same, probably not

Page 179

the same variables. What's a heavy cocaine use day? No binges are okay? Probably not. Marsden did a trial in the UK where they used a clinically significant -- I can't remember -- the RCI, and then determined that people who reduced their cocaine use 15 days a month was a success. But there are still people using 15 days a month. Is that really a success? It just gets so complicated in our field.

DR. STRAIN: Thanks. Other questions or thoughts?

DR. FALK: I just had a clarification question. Sorry. This is Dan, NIAAA. When you said urines are quantifiable, I'm really naive what that means exactly.

DR. CARROLL: Our urines can be yes or no, so that tells us something about how specified it is, but you can also -- if you get adequate urines and you do quantitative analyses, you can get I think a nice reading in terms of the levels going up or the levels going down, those kinds of things. It sounds great, but it tends to be really

Page 180

complicated. And again, it's very expensive to do it. But we can actually see.

DR. FALK: Do they tend to do that in most trials? Is it very rare to do that kind of quantitative -- in terms of the data that you're going to find that you have when you look across trials.

DR. CARROLL: It was very popular for a while. Kenzie, you know better.

DR. PRESTON: I don't think that quantitative analysis of urine results is very common at all. People have done it and then abandoned it. So I think the cost benefit hasn't worked out very well.

DR. FALK: Okay.

MALE SPEAKER: I think the trials where you've potentially toward a regulatory filing, you'd probably find that it's been done in all of those. That's the minority of the studies. Most are done for publication.

DR. FALK: I guess where I'm going, I was wondering what you can do with that quantifiable

Page 181

data. Can it become more like a continuous variable, or is it sort of just more like -- I just wonder -- I don't know; it's a kind of rhetorical question -- what more can be done with that than just having a yes/no.

DR. STRAIN: Kenzie, do you want to --

DR. PRESTON: One of the things my lab did is -- what we found is a huge discrepancy between positive urines and self-report, and so we did quantitative analysis, and we saw that some people's concentrations were going down. But because the cut-offs were positive or pretty low, and they're really designed to catch everybody, not to distinguish level of use, that we developed an algorithm to try to tell when people's positive was due to some recent use or carryover because we were testing three times a week. And in fact, we were able to normalize it so that what appeared to be the real rate of use was somewhere in between urine positives and self-report.

You still have -- and I will show some concentration data tomorrow. But the range for

Page 182

cocaine -- and we use a cocaine metabolite -- is huge, so you have to -- if you just went with concentration, you have to do transformation of the data because you can end up in the hundreds of thousands nanograms per mL, where the cut-off is 300.

DR. STRAIN: Dave, did you have a --

DR. McCANN: In listening to Raye and Dan's presentation, and then these most recent ones, in a way, we're trying to learn from the alcohol field and use what the successes have been there to help guide us. But there are some real important differences that makes it difficult.

One of them, after having set through, is the active meeting group, been meeting for seven years now -- through seven years of meetings twice a year. One of the nice things is, you guys have effective medications. When you start asking how sensitive is this endpoint, you've got data sets with effective meds. And for cocaine addiction, we just don't have that. That's a huge difference and a huge challenge.

Page 183

But the other thing, when you consider a non-abstinence endpoint, like no heavy drinking days, that's something that now has been endorsed by the FDA. Patients may go to their doctors and say, I don't want to quit drinking, but I want to reduce my drinking, and it's okay to have that as the goal. But for cocaine addiction, where anytime somebody buys cocaine on the street, it could be anything. A dose could kill them acutely. It's a very different situation.

This is a question that I would throw out for the group. Can anybody imagine simply reduced use as being something that a physician would say people should work toward as opposed to, you need to try to quit, and if you can't quit, if you can reduce the use, you reduce it as much. But it seems like the goal of quitting is always going to be there for the physician and probably in our clinical trials, too.

So if that's the case, it doesn't mean that a non-abstinence endpoint won't be useful if in the process of trying to quit, we find that a

Page 184

medication is able to substantially reduce use. Potentially that could be useful. But I think we need to think about the fact -- am I right in saying it's a fact that it's always going to be the goal to quit? And if that's a big difference, we need to think about that in terms of how acceptable is an endpoint.

DR. STRAIN: Thanks. Great. Connie and then Ivan.

DR. WEISNER: Just a couple of quick points. One is, on the health policy side, if you're studying treatment populations, that can be charged. Drug testing can be charged. It's like for diabetes, you need to have your insulin tested. So there's no reason for treatment programs to not be doing your end test. Many have to anyway because they have to give the stuff back to the employer.

Also again, as Dan and Raye and the group really know, we're not always so sure about how patients are answering drinking questions either. So when we do our expert in primary care, we just

Page 185

really have to have the drinking pictures on the wall because people say something differently if they really see what is a drink, what does a drink look like and so forth. So again, isn't it just the -- there's low fuzziness there, so I don't think this research should stop because of that.

The last thing I would say, we're still -- even on the alcohol side -- not looking at reduced use by looking at treatments that reach reduced use. We have abstinence-based treatments, and then we look at reduced use. You know what we might really -- there have been a few controlled drinking studies in the past and so forth, but for all of these things, you're right. We have to get buy-in from physicians and other clinicians to look at that. But until we really do that and really develop some treatments that would be focused on that, we can't answer some of these questions, but it shouldn't stop us.

DR. STRAIN: Thanks. Ivan?

DR. MONTOYA: I have a question. I'd like to know your opinion about -- the DSM-V has three

Page 186

categories of substance-use disorders: mild, moderate, or severe.

DR. STRAIN: Can you try -- is he getting picked up the mic?

DR. MONTOYA: Do you have some thoughts about using those categories as treatment outcomes and maybe trying in the future to test those categories, if they mean anything? Because they were just taken without any valuation, but they are now part of the DSM-V.

Any one of you? Kathy, you have all this analysis with different outcomes, but the DSM-V outcome is --

DR. CARROLL: As a measure of severity. Those weren't exactly empirically based.

DR. MONTOYA: Right.

DR. CARROLL: So we can go do that. And again, it's generating the database. It was in fashion for a while to repeat the skid at the end of treatment to see how much movement there was in time, but we didn't do that consistently enough in this to do it. We sort of reinstated that. So we

Page 187

intend to do it, but we can't give you an answer yet about how that maps out. By the time we get the answer to that, they'll switch to DSM-VI, I'm sure.

It took us 20 years of trials to get the 400 people that were clearly described with skids, and we had a year of data on functioning and healthcare utilization. It takes a long time to do it. I'm not saying not do it. I was pointing out the complications.

DR. STRAIN: Let me comment. But I think it's an interesting idea going forward because it's out there, and it's widely available, and could it be used as a -- as we think about what should be developed, it's a really appropriate and useful question.

Raye?

DR. LITTEN: Just a question. In our alcohol trials, we get people coming in that have a goal of abstinence and a goal of just cutting back. We sort of get a mixture. I was just wondering in your trials, does everybody come in for a goal of

Page 188

abstinence or do some say I just want to cut back? I was just wondering about that. It would seem to me it might be -- just accept those that really have a goal of abstinence.

DR. CARROLL: Right. Those 4 people, that would be great. It's not a lot. One of the advantages of the data set is that we actually had those data. And I think it's -- most people say -- it's 70 percent maybe; I would have to look -- say that their intention and their expectation would be to do that. In turns out to be utterly unrelated to how they actually do. So their stated goal and expectation prior to participating in one of our trials -- it predicts -- people are pushed in for all kinds of -- the heterogeneity is really a complicated thing.

DR. STRAIN: Goes back to Connie's point.

Kelly, you looked at -- was it the project COMBINE that you looked at pre-treatment drinking goal? Yes. And so it did predict, right?

DR. DUNN: It did, yes. Kind of the outcome

Page 189

of that paper was that it may be better within the alcohol field to evaluate outcomes in the context of the person's original goal, whether their goal was to maintain abstinence or to just reduce their use.

DR. STRAIN: Yes. So another stratification variable, Dave.

DR. McCANN: I just thought of that.

DR. STRAIN: Yes. We're up to four.

DR. McCANN: You're reading my mind.

DR. STRAIN: Well, I'm a psychiatrist.

I have a question. I know that we're coming up against lunch. It strikes me to go back -- if I follow -- and I'm going to betray my naivete about something here. But if I follow the logic of what was done in the alcohol field, basically, you use the drink as the measure of consequences to establish the value of heavy drinking days as the outcome measure of relevance. Correct?

That's critical, just to summarize, because the fact of the matter is, the drink in a short interval doesn't show a significant effect, but

Page 190

heavy drinking days does. But your logic is that you've got too short a window in the studies for the drink to show that.

Hey. The field has accepted it. You've got a paper that's probably getting cited hundreds of times, right?

MALE SPEAKER: I think Connie brought up a good point, though. Maybe it's not just the duration. I don't know if we really know why the consequences are not significant. There could be a variety of reasons.

DR. STRAIN: That was the preface to the question, which is does the stimulant field need a drink?

MALE SPEAKER: [Inaudible - off mic.]

DR. CARROLL: The short of it is there are short inventory problems. You have the SIP [ph] that we use.

MALE SPEAKER: I was going to talk with the Lily folks about how IMBIBE was different than the SIP, because the SIP was just kind of the shortened version of the drink, basically, right. So how

Page 191

does the IMBIBE -- is it just a different empirical process coming up with that as opposed to the shortened, abbreviated version? Yes, we don't have the consequences measure in stimulant trials.

I think to follow up on your point, though, Eric, it also was largely based on the definition of heavy drinking to begin with, of having consequences, right? Dan, Raye, you guys weren't the first ones to come up with heavy drinking day, right? That there was an already established definition for that meant something. And where that cut-off criteria came from, I was going to talk to you guys later about this. I'm a little ignorant about that aspect of it; where did that come from, and how is that somewhere we can use that in a cocaine trial. I don't know if that's impossible. Because it's largely dependent on how the outcome was validated because there was a definition of what's heavy drinking versus just abstinence.

DR. LITTEN: I don't know. Maybe Raye -- I wasn't around for that at all. I know 4 or 5

Page 192

drinks is probably, what, could be enough to get somebody drunk enough to start experiencing acute consequences perhaps. But I don't know if all the drink items would be -- are necessarily sensitive -- had that in mind that somebody had to be drunk in order to get the consequence. I think probably even a few drinks maybe could trip some of the consequences. I'm not sure what the threshold is.

MALE SPEAKER: Well, yes. That goes back. It's sort of a long history how it sort of developed. Even the late '60s, people were suggesting a drink was around 5 drinks or 6 drinks. I think Tom Vaver [ph], back in the late '80s -- I think Hank Kranzler was part of that, too -- pointed out the difference that seems to be a good cut-off for problems.

Martha Sanchez, which Celia actually quoted to, looked at a couple studies and found those who had the 5 drinks had problems, those that didn't.

Then we started validating this 4 or 5 in recent analysis that I showed today, and it just

Page 193

seemed to work out well. And I know Celia's been looking at this, too.
DR. WINCHELL: So I'll just add that different cut points were explored. There are some unpublished analyses, regrettably unpublished analyses, that involved sort of diving into the MATCH data set and swimming around, and explored various responder definitions, either based on absolute cut-off, different levels, or percent days abstinent, all kinds of different approaches, and emerged with support for this particular endpoint.
Similarly, the analysis of the NAS data set explored a whole lot of different cut points, and kind of converged on this point, that we found those two convergent lines of evidence supported recommendations that we started making a few years ago, and then additional lines of evidence continued to support that from the managed care data set and from the NESARC.
DR. STRAIN: Joanne, were you stretching?
DR. FERTIG: No, I have a question.
DR. STRAIN: Okay. Yes, then we'll break.

Page 194

DR. FERTIG: This is going to reveal my naivete, and it's for Celia. Is it really likely or possible that the FDA would approve anything other than an abstinence outcome for an illegal substance?
DR. WINCHELL: Well, I'm going to echo what my boss said earlier today, which is --
DR. STRAIN: It's always good to echo your boss.
DR. WINCHELL: -- we are here to listen to what the field has to say. We also think that the field needs to be aware of what providers, payers, and policymakers think as well because, otherwise, you just sort of wind up preaching only to the choir. But we are here to learn from you.
DR. STRAIN: On that note, why don't we break for lunch, and let's reconvene at about 1:25, which means it will be 1:30. But if I say 1:30, it will be 1:35. And lunch is back where breakfast was, down this way. Don't any food in the hallway. I remind you again.
(Laughter.)

Page 195

DR. STRAIN: Thanks, everyone. And thanks to all the presenters this morning for a great session.
(Whereupon, a lunch recess was taken.)

Page 196

A F T E R N O O N S E S S I O N
DR. STRAIN: So we're going to get started, and I'm going to ask Dave McCann to come up. He's going to be moderating this session, which will now go until 3:00. And the title of this session is Prior and Ongoing Efforts to Evaluate Clinical Benefit in Stimulant Trials, Based on Past Studies. So David, I turn it over to you. Thanks.
Moderator – David McCann
DR. McCANN: We're almost all back here I guess. Within NIDA, we've been working, in a stimulated work, evaluating potential endpoints for a number of years, going back at least four or five years. We had grant supplements available in 2011 and gave out three of those to work on existing data sets to really try to go beyond abstinence as an endpoint.
In the discussions we've had with the FDA folks, abstinence, especially end-of-study abstinence, being clean at the end of a trial, is accepted as an endpoint. Now, we may still debate how long the period of abstinence needs to be and

Page 197

how long the trial needs to be, but for a certain period abstinence, we don't need to prove that it's beneficial. It's accepted that it's beneficial. It's when we get to something other than abstinence, intermittent use or reduced use endpoint, that the message we've gotten from Celia and others at the FDA is we're open to hearing about it, but we're data driven. You need to bring us data that shows us that something other than abstinence has a clinically meaningful benefit.

So that's the message that we tried to get out to the field over the past few years, and we've tried to support work to generate data that could drive the FDA, pharmaceutical companies, payer, to be convinced that something other than long-term abstinence is an acceptable endpoint.

So out of the four presenters that are coming here, three have worked through grant funding, and Ivan Montoya along with Shou-Hua Li, who's the statistician in our division, spent some time looking at our previous contract supported medication trials to try and pull out data that

Page 198

might show clinically meaningful benefit to something other than long-term abstinence.

So we'll start out with George Woody.

Presentation - George Woody

DR. WOODY: Thanks for inviting me. I'm going to present data from one of our NIDA supplements. Unfortunately, I didn't put the grant number or anything up there. It was a supplement to our CTN node, the Delaware Valley Node, one of the supplements that was just mentioned. Then at the end, I'm going to put some data from a study of naltrexone that we did in Iceland, very quickly.

The goal of our supplement was to do secondary analysis of data from the NIDA Cocaine Collaborative Psychotherapy Study. This was done a number of years ago. Paul Crits-Christoph was the PI, and the treatments were group drug counseling, counseling plus cognitive therapy, counseling plus supportive expressive therapy, or counseling alone.

The secondary analyses were guided by a concept paper by David and Li, who demonstrated that the number of weeks of abstinence during and

Page 199

after treatment shows differences in outcome between treatments, even when previous analyses found their differences. So the goal of this study was to evaluate the merits of different methods of measuring end of study -- measuring outcomes using different ways to look at cocaine use, both during this study and measures of overall functioning at follow-up.

The design of the parent study, it was six months of counseling with or without psychotherapy treatment, a maximum of 24 group sessions. It was group plus individual. Everybody got 12-step oriented group therapy, and some got additional individual therapy. The counselors were highly screened. We screened them. I participated in this study, trained them in drug counseling according to the manual that's on the NIDA website. Almost 500 patients were randomized. We assessed drug use and overall functioning using the ASI during treatment and up to 12 months after randomization, and urines and self-report were used to assess drug use.

Page 200

Here's the patient's sample, 50/50 occasion, African American, most using crack; about a third alcohol dependent; about a third with cocaine-induced mood disorder. This is the attrition that we had follow-up data of drug use, either urine tests or self-reports on 85-90 percent of the subjects. But they kept a little less than half of their psychotherapy sessions, which is pretty much like what we've gotten when we did psychotherapy in the methadone program.

Here's our outcome, a surprising outcome. As you see, the hypothesis, adding the extra therapy would help, as we found out with a study in the methadone program. But here, it was the drug counseling that did the best. But everybody got better. And actually surprising was the group drug counseling alone. Excuse me. That's SE therapy. But the group drug counseling alone did pretty well, too. Percent of patients achieving 3 or more months of abstinence; there you see it. Individual drug counseling, again, had more in that group.

The questions were what measures during

Page 201

treatment cocaine use best predict end of treatment functional outcomes, and drug-use outcomes at follow-up, and functional outcomes at follow-up? We were guided by what David had done. Basically, what you see here, in this series of analyses, are cocaine outcomes predicting 12-month follow-up functioning. The significance, here you see psychiatric -- and all these are fairly low. When they're significant, they're fairly low effect sizes so to speak.

But there you see psych was significant. But the most consistent one, actually, was legal, which is why I asked the question about legal outcome earlier, which is not surprising because cocaine's illegal. So those who are using more would be more likely to have legal problems. Also, the psych difference is not surprising because cocaine is psychotoxic. It produces a lot of psychiatric symptoms, so less use is usually associated with less psychiatric.

Here we looked at predicting 12-month drug-use outcomes, and basically what we found was

Page 202

drug use during treatment-predicted drug use had outcomes, which is pretty much what a lot of people had found. That's what it was.

Then we used some of the methods that David McCann and Li developed. We looked at end-of-study abstinence and number of -- 2 weeks here was the threshold. And then we looked at number of beyond threshold weeks of abstinence to see if that made any difference, and basically it didn't.

If we looked at within-treatment cocaine outcomes, predicting 12-month drug-use outcomes, basically what you found was within treatment, no matter how you looked at it. Any one of those parameters within use predicted later use. And that was significant in just about all the measures. And if we looked at functional outcomes as they related to these, basically there was nothing predicted. There was just one significance, and that was with employment. But there were so many measures, that could have just been an incidental finding.

What we concluded from that, from this

Page 203

study, within study, cocaine use measures moderately associated with cocaine-injecting drug use at 12 months. The abstinence measures actually looked best here. And within-study cocaine use wasn't associated with function at 1-year follow-up. And end-of-study abstinence and weeks beyond the threshold, established by David and Li, associated with cocaine abstinence measures at 1-year follow-up.

So we basically didn't find any gradation, much of a gradation here. From this sample, it looked like the drug use is episodic for these cocaine users. There were relatively few intermittent users, mostly continuous users for abstainers. Some people seemed to functional relatively well despite drug use, and we really couldn't summarize outcome as one successful index in these data. A lot of limitations, a limited number of outcome measures.

We really relied on the ASI plus urine and/or self-report. There was a group of complete abstainers, so there was a somewhat restricted

Page 204

range of outcomes, and patients with psychiatric comorbidities requiring psychotropic medications were excluded.

Just a few comments on this Iceland study with naltrexone. If anybody's never been to Iceland, I would really recommend it as a place to visit. It's a very, very interesting place. And we wanted to see if we could see not excessively replication I guess, but a signal that naltrexone worked for amphetamine users because at the time we did this, amphetamine use was a big problem up there.

The study that we based on was something that was published from Sweden in the American Journal of Psychiatry. And there they found that -- and this was oral naltrexone with or without -- everybody in drug counseling, with or without oral naltrexone, and there you found a significant less positive urines in the naltrexone group.

This study -- and we're working on a paper writing this Iceland study up. You really

Page 205

can't -- we sort of thought of this, well, it's the same group. It's a Scandinavian group. They're genetically similar. But the patients in this population were real -- there's a different sample.

These patients were very carefully selected, people who were only amphetamine dependent. They rejected three-quarters of the people that they screened. They had no alcohol dependence. No dependence on any other drug except nicotine, and they had to demonstrate two weeks worth of negative urines before being randomized into this study. So they really focused down on pure amphetamine dependence.

In Iceland, our population was very different, as you'll see in a minute. Also, from what I've seen, they've got the best treatment programs that I've ever seen. They have centralized addiction treatment. Everybody gets hospitalized before starting treatment. Good access treatment's free. This is a hospital. The population endorses the disease concept. They feel that addiction is a health problem, not a legal

Page 206

problem. The staff, to be certified as a counselor, you have to pass a national exam after two years of training. And they really use evidence-based practice.

That's just a summary. Detox first, 7 to 10 days in the hospital. And then they either go to residential or directly to IOP with follow-up outpatient.

We randomized 100 people primarily seeking problems for amphetamine dependence, randomized to Vivitrol or Vivitrol placebo. Alkerme's provided the medication free of charge, and we're very happy about that, stratified by gender and IV status. And they were randomized before they went to the outpatient treatment because remember, some went to residential and then outpatient, and others went directly from Vogur to outpatient.

The population is really different than the population in the Swedish study. They were all dependent on amphetamines, and they were averaging using a little over 18 days a month, a little bit more than the Swedish group. But look at that.

Page 207

Three-quarters are alcohol dependent, cannabis dependence, cocaine. There was a lot of other drug use among this group. This was a heavy drug-using group. Very little opioid dependence. That was an exclusion criteria, but there's very little of that in Iceland.

This was retention in the study treatments, as you see, and that is getting the Vivitrol injections. And here, we've got urines. We've gotten a total of 2400 urines on this entire sample. We got a little over 1200, a little bit more than half. And look. Of the 1247 urines, only 53 were positive for anything. A statistician looked at that. Kevin Lynch didn't believe they were amphetamines addicts because that's a very low rate. But then, if you impute missing urines as positive, that's what you get there.

So it looked like about half the patients were doing pretty well, and those were primarily the ones that stayed in treatment. And this is the percent of drug-negative urines, amphetamines. Benzos were the most common, but it was very low

Page 208

across the board.

So really, we didn't see anything. We didn't see a signal of a naltrexone effect. There could very -- if we looked at a more restrictive sample, perhaps, like the Swedish did maybe, we would have seen it. We have not analyzed the data as per -- McCann and Li will be doing that.

That's just a little overview. We're going to have trouble analyzing the data according to Li because we are missing a lot of follow-up data on people that dropped out of treatment. A contributor to that is that the Icelandic IRB would not approve patient payments to come back for treatment. They said, look, they're getting treatment free. It's a national policy. Because we wanted to put patient payments in to get them to come back, but it couldn't get through their IRB if we put that in there.

We do have some follow-up data from telephones and self-report and all that kind of stuff, but we're going to have issues with that. But that's something we're going to be working on.

Page 209

So that's what I have to present. Thanks.
DR. McCANN: Just a couple of questions (inaudible - off mic.] I can start it off. For the Vivitrol study in Iceland, I actually saw those data presented a year and a half ago. I got to apologize for my memory. So I need to ask, did you look for any correlation between amount of amphetamine use in these subjects and some other endpoint to look --
DR. WOODY: Yes. We will be looking at that. Craving went down. What I have is craving went down, again, in the people that stayed in treatment. We haven't drilled down on that. One of the things that's been a delay is that we have all the data, but they were originally entered into the database in Iceland, and there were issues getting it correctly moved over to us. So we're a little slow in that.
DR. McCANN: Let me encourage others, if you have any questions about the first part of the presentation on the cocaine data and looking for evidence of a meaningful benefit to reduce use,

Page 210

this is an opportunity. I thought it would be good to go through some of the presentations, even if we didn't have clear findings to support a clinical benefit to something other than abstinence, so that we could ask exactly what was looked at. I encourage you to suggest other ways of looking at the data that we might consider.
DR. MOELLER: I have a question over here. Gerry Moeller. Did you look at potential ER visits? We were talking about this earlier. This is an aging population for the cocaine use. I see a lot of patients on the inside of the hospital who've had cardiovascular complications with cocaine use. And each episode of use is potentially a risk factor for those complications. So did you look at that?
DR. WOODY: We looked at other treatments received. Some patients did receive treatment outside the study. ER visits are probably in there, but I'd have to double-check. I don't remember. I doubt if there are too many ER visits. There were readmissions, and the readmissions would

Page 211

come to Vogur Hospital.
We have that. But again, we're sort of slow in getting this out because of some issues with the translation of the databases, not just from Icelandic to the U.S., but the guy that did it in Iceland was like a self-taught data entry guy. And he got it all in there, but it was a little hard for us to make sure it was in our format.
Yes?
MALE SPEAKER: For the first two trials, it looked like you had a treatment effect. Did you compare the sensitivity of all the different outcomes among -- like effect size for all the different outcomes to see if they had different effect sizes, basically.
DR. WOODY: I'm not sure. It looked like -- with the cocaine psychotherapy study, it went -- everybody got better. Everybody went down. You did have that significant difference between -- and the others.
MALE SPEAKER: On that outcome, you had the Y-axis -- I forget what the outcome is. Maybe it

Page 212

was percent negative urines. I forget what the outcome -- the Y-axis. But you could reproduce that same graph with a bunch of different outcome measures and see if it kind of replicate -- well, maybe not that same graph, but you could see if the treatment effects were similar if you used different outcomes. It could be a way to sort of just test the sensitivity of different outcomes.
DR. WOODY: Yes. Okay. Thanks.
DR. STRAIN: Thank you.
(Applause.)
DR. McCANN: Our next speaker, Brian Kiluk, did some work with Kathleen Carroll, and we're really excited to see what the results of some of the analyses were related to the cocaine use.
Presentation - Brian Kiluk
DR. KILUK: Hello, everybody. We'll be talking about the work that I've done with Kathy as a result of the NIDA supplement that we received. Kathy set up some of this in her talk, at least discussing the indicators that we decided to look at, as well as the challenges with defining each of

Page 213

those indicators. So I'm not going to go over those.

Basically, what we did was we evaluated these 15 candidate cocaine-use treatment outcome measures after pooling the data across five completed randomized trials for cocaine dependence. We evaluated them according to criteria in terms of sensitivity to medication effects, sensitivity to the behavioral therapy effects, correlations or relationships with post-treatment cocaine use, so during the follow-up period, as well as to measures of general functioning, which Kathy kind of talked about, our straw man measure of general functioning. And this was published last year. The first portion of this, the data that I'm going to go over here was published in drug and alcohol dependence.

These are just a list of the five trials that were included, that were all completed, studies that have been published on. The last one is currently under review, but studies that have been conducted over the last 15-plus years.

Page 214

The nice thing about all these trials is, since it's within the same research group, we have common assessment batteries and all and a common assessment approach. Almost all of them, 4 out of the 5, included a 12-week treatment phase. One of the trials was only an 8-week study. That was the computerized CBT trial.

Virtually, all looked at some sort of behavioral therapies. The behavioral therapies were manual guided with independent fidelity ratings, medications, were placebo controlled. Urine testing ranged from 1 time a week to 3 times a week, depending on the study and the setting.

All of the trials included a follow-up period that assessed substance use as well as other measures up to 1 year. So 4 out of the 5 trials included a 12-month follow-up period, where the one trial that was only an 8-week study only included a 6-month follow-up. But we essentially have data for most -- up to 15 months from the time of randomization.

Our primary measure for substance use,

Page 215

self-reported at least, was the Substance Use Calendar, which is essentially the timeline follow-back, a calendar-based method. For that, we can get a day-by-day frequency of cocaine use during the entire study. I think some of the earlier studies, we tried to gather information on the quantity of cocaine use, but by and large, we didn't have that across all. So we're really just looking at the frequency of cocaine use.

As Kathy mentioned, there was some discrepancy with urines, although that was pretty low. On average, I think it came out to around 13 percent, although largely, that number was driven by some of the earlier studies where we used a laboratory for testing urine. So rather than getting the instant result, which decreased the rate of discrepancy, some of the earlier studies had a bigger discrepancy. But overall, a fairly low rate of discrepancy between self-reported cocaine use and urine results. So that would be a positive urine with a denial of self-reported cocaine use in the three days prior to that urine.

Page 216

We have good rates of follow-up, greater than 80 percent on the intention-to-treat sample across the studies. One of the common assessments that we used across all the studies was the ASI, so we had some nice data there on the ASI.

In terms of during the follow-up period, we were able to calculate mean days of cocaine use at each of the follow-up time points, using a substance-use calendar. One of the additional measures we chose to look at was complete abstinence through the entire follow-up period. So again, this would be based on self-report as well as at least a clean urine at the moment of the assessment.

Then we also wanted to look at some measure of global functioning or good functioning, what we termed initially -- there are a couple of different labels for it, but it ended up being called our good outcome or good enough outcome. So this was based on looking at the ASI. And rather than evaluating the composite scores from the ASI, we chose to use a bit more of a patient-reported, I

Page 217

guess, outcome, just looking at the days of problems across each of the domains on the ASI.

So we selected the ones that we thought might be most salient, which would be days of employment problems, days of legal problems, and days of psychiatric problems. So if everybody had reported zero days across all of those, as well as had zero days of cocaine use, if they met that criteria, we considered that a good outcome or good enough outcome.

I'm going to breeze through some of these next few slides. This is just to give an overview of what the studies included in the data set look like. Study 1 was a 3 x 2, included clinical management, 12-step facilitation and CBT, as well as either disulfiram or no med, with 121 cocaine- and alcohol-dependent patients.

Study 2 was CBT versus interpersonal therapy, and included again, disulfiram versus placebo with 121 cocaine-only dependent patients. So these didn't also have comorbid alcohol. So this was again, a 12-week trial.

Page 218

Study 3 was another 12-week trial. This one evaluated treatment as usual versus treatment as usual plus 12-step facilitation, again, with either a disulfiram or placebo. It was 112 cocaine dependent. And these were patients on methadone maintenance, so this sample is a little bit different than some of the rest, but still, cocaine-dependent and actively-using participants.

Study 4 was our pilot trial of the computerized CBT. This was just a two-group study looking at treatment as usual versus treatment as usual plus the computerized CBT. This trial was at an outpatient substance-use facility and took sort of all-comers, any kind of drug use. And for the current analyses, we only included those that were primary cocaine users. So it's 38 out of the 78 participants that are included in this pooled data. The last study was a CBT versus CBT-plus contingency management, again, with a disulfiram placebo platform on that. That's included in 99.

Just a general overview of the trials in terms of what the sample looked like. Down here,

Page 219

we kind of have the total, so we ended up with 434 participants across the trials that had some data during the 12-month follow-up period. About a third were female. Roughly half, ethnic minorities. Mean days of cocaine use at the time of baseline -- so the month prior to starting the study -- was 13 days out of the last 28. About half of them were not working, and 16 percent were referred by the criminal justice system, although there is some variability across the studies here.

In terms of just general outcomes across the trials, just so everybody gets a sense of what the outcomes look like in this data, we see the rates of percent of cocaine-free urines across the study. They're generally pretty consistent, although the 12-step disulfiram study, which was in the methadone maintain sample, which seems to be a bit of a outlier here -- but the rest are fairly consistent, percent days absent and during treatment range was in the 70's to 80's, although except for that third study.

Percent completely abstinent during

Page 220

treatment, we see it's not zero. So there are a small proportion of people that were completely abstinent during the treatment period. And then here's data on our 6-month follow-up, so the percentage of people that achieved that good-outcome criteria. Initially, when we came up with the variable, we weren't sure if anybody would actually meet that criteria. So it was kind of surprising that we did have some people that actually achieved that. So they were not using cocaine, and they reported no problems in legal employment or psychiatric areas.

So what we did is we took these 15 candidate indicators, and then evaluated them to whether they were able to detect medication effects. Since the disulfiram was the main medication used across the trials, we were able to see whether they were sensitive to detect effects across these indicators.

I realize there's a lot of data on here, so I've just tried to highlight the things or direct your attention to those things that are

Page 221

highlighted. The indicators that seemed to come out here, at least be able to detect the effect of disulfiram in this case, we have days retained in the treatment protocol, maximum consecutive days of abstinence, so several of the continuous indicators that have been discussed already, percent days abstinent.

Some of our included dichotomous outcomes that weren't detecting effects, people that achieved at least a certain number of weeks of abstinence during the course of the study, and we have some reduction measures here, too, achieving at least a 50 percent reduction versus a 75 percent reduction.

I would state that our reduction measure is fairly crude because, again, we didn't really have quantity, so essentially we were looking at the days of use. So we're comparing the days of use prior to starting the study to the days of use in the last month of the study. And that achieved at least a 50 percent reduction of a 75 percent reduction.

Page 222

So we have some indicators that can detect medication effects. Here it is looking detection of effects of behavioral therapies, again, some of the same indicators come up here, the continuous indicators, percent days of abstinent, proportion of cocaine negative urine samples here. And then some of our dichotomous indicators, particularly the 3-plus weeks of abstinence, so achieving at least 3 weeks of abstinence, I was able to detect effects of the behavioral therapies.

Our good outcome here actually was kind of lit up on this one as well, so people at the end of the trial achieved that criteria for good outcome. That differentiated across the behavioral therapies, too.

Another slide with a ton of data on it, and I'll just try to summarize this. This is an important one, which is where we -- again, the 15 outcome indicators here, evaluating them according to their relationship with post-treatment cocaine use. So the days of use at each of the follow-up points, 1, 3, 6, and 12 months.

Page 223

Relationships with whether individuals were completely abstinent during the follow-up period, and then the relationship with whether people achieved a good outcome at each of the follow-up points.

The shadings represent significance, so anything that's not shaded was a non-significant correlation. The lighter gray was considered a lower -- it's significant but a magnitude of correlation less than .3, and then the darker grays were magnitudes greater than .3, so just to give a highlight of some of the stronger relationships. You can see the darker areas here. Again, the continuous variables seem to come out at least predicting or related to follow-up drug use, cocaine use during the follow-up period.

Our indicator of 3-plus weeks of abstinence seems to show that as well. Interestingly, the ones that weren't showing up very much was the days retained in the treatment protocol wasn't related to much post-treatment cocaine use or good functioning, as well as some of our reduction

Page 224

measures. And again, our reduction measures were pretty crude, so we kind of had a -- they were kind of behind the curve, behind the ball on that one.

This just summarizes what we've found in these evaluations. So again, thinking about our criteria of sensitivity to disulfiram effects, or a medication effect. Sensitivity to the behavioral therapies; relationship with post-treatment cocaine use; and then relationship to general functioning during the post-treatment period. An X indicates that that criteria was met.

I've highlight the ones which I think are important, and red indicates ones that just didn't perform very well in our analyses, in our data set. So days retained in the treatment protocol, achieving at least one week of abstinence during the treatment period; or again, as Kathy mentioned in her talk, being completely abstinent during the treatment period was, again, sort of a surprising one that wasn't associated or didn't detect any of these effects, as well as the 50 percent reduction.

Ones that seemed pretty strong was the

Page 225

abstinence in the last two weeks of treatment. Again, that was one that we evaluated, which was of interest, and this dichotomous outcome of achieving at least three or more weeks of abstinence. And that three weeks is at any time during the treatment period, not just at the end.

As a follow-up to this, one thing that we decided to look at was thinking about the relationship between within-treatment cocaine use and follow-up, functional outcomes. So as opposed to just looking at this dichotomous of zero days of problems across several areas on the ASI, we wanted to be able to look at a more continuous way of looking at those by using some latent growth curve models to see if within-treatment cocaine use had a relationship with the change in these problems over time during the follow-up period. And this came out last year in JCCP, for those who are interested.

Just to describe it, our global problems construct, again, we created this latent measure of global problems, again, using the item from the

Page 226

ASI, which was the patient reported days of problems across each of the domains on the ASI, except for cocaine or alcohol use, so the non-drug using domains. Several steps we utilized here to confirm the construct as well as evaluate using conditional latent growth curve to examine the association between within-treatment cocaine use and to follow-up global problems over time.

This is our spider-web looking model, which I realize probably isn't very useful to anyone outside of this room or to some people who are in this room, but I'll walk you through it because I think it is important.

Across the top here is our latent measure of global problems at each of the follow-up time points. Again, these are indicated by the six domains on the ASI days of problems. I think the important thing to note about this is that rather than just evaluating each of these domains separately, this is accounting for some of the variability. So some people might come in with more problems in the legal aspect versus some that

Page 227

might come in with more problems in the employment.

So these latent constructs are accounting for that. And that's taken it into account rather than looking at each one individually, where you might have people that don't have any employment problems, this accounts for that.

Our within-treatment cocaine use measure here, we actually looked at this two different ways, one of which was a latent construct, which was indicated by three continuous measures: the percentage of days abstinent, the maximum consecutive days abstinent, and the percentage of -- this was positive cocaine urines.

Just to highlight, this was our within-treatment cocaine use indicator. What we found is that was related to post-treatment cocaine use, which was significant, which is great, which is what we kind of saw in some of the earlier stuff. The post-treatment cocaine use, so the cocaine use during the follow-up period was related to the overall average problems during the 12-month follow-up period. So same period of time here.

Page 228

Cocaine use is associated with more problems.

But interestingly, then, the within-treatment cocaine use was actually also related to this average level of global problems during the follow-up period, so greater abstinence was associated with fewer problems, and that was significant. And this is also controlling for the baseline level of problems, so it's not just that people have the same problems coming in.

This was actually replicated with the dichotomous outcome of just 21, 3-plus weeks of abstinence, so we looked at that, the same essential pattern here, similar coefficients, similar pattern of significance, suggesting that abstinence during the treatment period was associated with fewer global problems.

Now it wasn't related to the slope, and this is likely because there was very little change in the number of problems or the days of reported problems over the course of the 12-period. It was more so kind of a stable aspect. So we may have seen more changes during the treatment period that

Page 229

were just sort of maintained during the follow-up, but which is the reason why we're thinking it wasn't related to the slope, which wouldn't account for the change in problems. It was pretty stable.

To summarize so far, what we found from our data was that the existing widely used continuous measures seemed to be consistent predictors of cocaine use and general functioning during the follow-up period. Again, the positive ones were percent days abstinence, the maximum days of continuous abstinence during the course of the study, percent positive urines. We've also looked at it as percent negative, which is what we're using more recently. And then days of abstinence in the last 2 weeks also seem to be pretty positive as well.

There was good performance of the urine measures and abstinence at the end of treatment. But again, it has to be taken into account with the availability of data and all the issues that Kathy described earlier with what do you do with missing urines or when people drop out, and you're looking

Page 230

at an end-of-treatment measure. How do you calculate that and how do you determine that? So there's lots of issues with that.

Measures that seemed to perform a bit poorer or at least in our data, again, were those reduction measures, which I mentioned. We don't really have a great way of looking at production, at least in our data. And that's something that we're still trying to figure out and work through now to see if we can come up with a better way to look at reduction over the course of time, rather than just looking at the frequency of days of use. But it's problematic.

Complete abstinence during treatment; this is another one that we're still trying to chase down to understand. Kathy gave the explanation earlier that maybe there is some benefit to having a use or a slip during the course of treatment that people learn from. We found that the people that were abstinent during the entire treatment period, that wasn't associated with their abstinence during follow-up.

Page 231

We looked at this, whether this was related to people who were coming into treatment who were referred by the legal system, so maybe they had some external pressure to be abstinent during treatment. But once that legal requirement was resolved, maybe they returned. That didn't come up so well. That was a good thought, but that didn't prove correct. So we're still trying to chase down that complete abstinence, but we do think there is some potential benefit to having a use during treatment that people learn from.

The end-of-treatment abstinence and the 3-plus weeks of abstinence seemed to be pretty consistent predictor. This might be a direction to kind of move in, and it's one thing that we're looking at more of to see if we can evaluate this in some other trials and other data sets to see if we're finding the same results because this could be a direction to go in. And then again, the higher levels of cocaine abstinence during treatment were associated with fewer problems.

So this is something -- we also wanted

Page 232

to -- a path we want to go down so that way we're meeting the requirement or the notion that there is some clinical benefit associated with these measures of abstinence. It's not just reduced use without any functional benefits. That's it.

(Applause.)

DR. McCANN: As the moderator, my threshold for falling behind is one complete presentation. So we're now 15 minutes behind. I think we can hold questions now and ask them during the time we've set aside for that for everyone. And I'm not going to bother taking the time to walk up there.

Dr. Shengan Lai, to go along with the previous Monty Python analogy, now for something completely different. This is not reevaluation of data from a treatment trial, but really a truly novel way to look at consequences of cocaine use.

Presentation - Shengan Lai

DR. LAI: Good afternoon. I'm new here. I don't know anything about the drug treatment. I got really scared. But I tell you, everything I did right now I owe Dr. Skolnick a lot because when

Page 233

I learned my data, I thought I got it trashed. Although the paper published, I didn't get anything out. But he said that this is a treasure. I report to you.

I could be wrong [indiscernible - mic interference] -- everybody knows. NIDA, the founding study -- the FDA has not approved any medications for treating cocaine addiction in humans. And FDA accepted the self-reported data for the heavy drinking days, but they don't accept the concept, the days of heavy cocaine use. Basically, the bar is much higher. They need a period of abstinence that last through the end of the trial.

FDA believes that cocaine use, behavior is only a surrogate indicator for risk of health and the behavior problems. That's why we need to target health risks associated -- potential disease markers. With NIDA to support it, we have enrolled and followed up 1500 African Americans in Baltimore for 14 years with a very low follow-up. The dropout rate is less than 3.5 percent per year.

Page 234

Everybody had the CTE. There's a contract in-house, the CT angiography, and we know every detail for the heart condition, the arteries.

Part of the study participants had an MRI to measure the left -- region of dysfunction. We have lab data. We have lots and lots of lab data. For example, recently we did an analysis, a [indiscernible] analysis. The [indiscernible] size is 1429. After adjustment for tool 13C, ECCEHA [ph], the cardiovascular risk, cocaine use was significantly associated with subclinical heart disease, but actually it was not. I would get these two papers out as soon as possible.

The objective of this study, the current study, we have three objectives. Number one, whether cocaine abstinence leads to less endothelial damage. Second is we explored whether cocaine abstinence retards coronary plaque progression. Probably the FDA likes that. And third, whether reduction in cocaine use leads to less endothelial damage in African Americans. Crack cocaine users with contract -- enhanced CT

Page 235

angiography confirmed less than 50 percent coronary stenosis. What do we want to see -- this is progression. Less than 50 percent stenosis is minor. When they cross the line above 50 percent, they call this significant stenosis. The cardiologist has to take over.

In March and April last year, we recruited 38 cocaine users. It took one year it took -- CCRC [ph] approved this study. This is criteria. We recruit everybody for ongoing study, and they have to confirm they are cocaine users. They have a urine test to confirm they did use cocaine. As clinical criteria, we don't want anybody with clinical heart disease. We don't want pregnant women in this study. We want people with good kidney function, to be approved of course.

We did an interview. I know time is running out. We did an interview, and like everybody, we have markers. We want the people to stop using cocaine. We pay them. We pay them to stop using cocaine. That's why CCRC will not say, look, you're going to create huge trouble for university.

Page 236

It took a year or two to get approved. It's worth it. We did the urine test using the Dip Card, and we did the CT angiography.

I show you the results. Right now, we have 74 people enrolled. The first two months, among 38 people, 22 people finished the study over a 6-month period, baseline to 6 months. We have data. We have data for the median of [indiscernible], and we have all other data, and I tell you want happened.

This is baseline characteristics of 22 African American chronic cocaine users. This is the data. This is baseline. Total abstinence from cocaine and reduction in cocaine use. Among 22 participants, 11 were abstinent from cocaine for 6 months, while 11 continued use. Part of the people reduced use. Only 2 people consistently used for cocaine for 6 months; only 2.

I show you the figure. Baseline, everybody had to use cocaine. Over time, the first month, second month, six months, the amount of cocaine is reduced because we pay. Then total abstinence from cocaine, and the reduction in Endothelin 1.

Page 237

Endothelin 1 is released from the injured endothelin cells. With other injury, endothelin would not come out that much.

So basically, because we have a follow-up study, we did the GEE analysis. I show you the table. Here's the table. [Inaudible - off microphone.]

It was very significant. When people quit from using cocaine, the marker dropped. This is follow-up study. And then we looked at the reduction in cocaine use; not complete quit but reduced use. You see the table. Again, you see the reduction [indiscernible]. The lower of cocaine use, the higher -- the higher reduction in cocaine use, the lower the marker.

Then we looked at the incidence of coronary plaque progression. That means from less than 50 percent to above 50 percent. Among those 22 people, 11 quit. Among those 11, [indiscernible] of them significant stenosis within 6 months. But among those 11, they still use cocaine, although some reduced significant stenosis.

Page 238

I sent this paper to a journal. One reviewer did not believe this. And I said, I will send you a picture. I will show you the table. Basically, this is two groups. One is they completely quit; no one used cocaine in 6 months, and none of them developed significant stenosis. Among those 11, 2 developed significant stenosis.

This is a picture. This is the first patient. [Inaudible - off mic.] This is at baseline. There's almost no stenosis. This is 6 months later, almost 70 percent of stenosis. As to where we got it, it's the same person.

(Laughter.)

DR. LAI: This is a problem. This is first patient, and then we go to second patient. This is the second patient. This patient had a calcification here. However, to deposit all [indiscernible] of artery -- will block any blood flow. But here, 6 months later, got 65 percent of the [indiscernible]. We refer this patient to the cardiologist. I think he's doing the [indiscernible] now.

Page 239

This image analyzes the [indiscernible], the doctor, Elliot Fishman. He's the chief cardiologist. Then I sent the data to Dr. Bluemke. He's the chief of radiology at the NIH clinical center. I said, "I want to use different approach to analyze the data." He's using new data. [Indiscernible]. Here is the LAD here. There's almost no stenosis here. It's gone. This is hard plaque; this is soft plaque. This is one patient. He didn't want to do a second one because he said it's too time consuming.

So the conclusions, I know it's limitations, but this study provides evidence that ET-1 could be used as a marker for cocaine abstinence or reduction in cocaine use. The findings of the study may also provide amazing new avenues of research in the fight against cocaine-induced premature coronary atherosclerosis.

When the people finished the 12-month study, I gave them a certificate. Some people cried. They said, "The study's over? Next week, where should we go?" I wonder whether or not this study

Page 240

can continue, I can continue producing these certificates. I want to give the certificate to those who completely quit.

This is a study. Everybody made some contribution. I have a special thanks to Dr. Skolnick because he told me this is good stuff. I did not know before. This study is supported by NIDA. Thank you so much.

(Applause.)

DR. McCANN: If you could stay there for just a couple questions. What you've presented is so new to the field. I want to give folks a chance to ask a couple questions, unless of course you were so clear that no one has any questions.

DR. STRAIN: How easy it to measure this biomarker?

DR. LAI: It's very easy.

DR. STRAIN: It's a simple test?

DR. LAI: Very simple blood test. Also, I have another marker. We are working on it. I have another marker, not the data yet. The hard part, I have to pay extra money to analyze the plaque

Page 241

volume because when the plaque grows, it's three dimensions then. It's not just cross-sectional. And also, I have to work with Dr. Bluemke. He analyzed the whole volume for the entire -- the artery trees.

DR. STRAIN: Do you have any sense of how -- I guess I'm struck by the fact that you're seeing -- if I followed you, I think you've got two subjects who you did with a 6-month interval and showed the plaque formation over that 6-month period. Is that right?

DR. LAI: Yes.

DR. STRAIN: But they've been using cocaine regularly for some time. Is that right?

DR. LAI: Yes.

DR. STRAIN: So it seems peculiar that at baseline, they had very little plaque accumulation. Did you just look at the LAD alone?

DR. LAI: No. Everybody had a little stenosis. We looked a -- you know, at the tree.

DR. STRAIN: Yes.

DR. LAI: The main thing is LAD.

Page 242

DR. STRAIN: LAD.

DR. LAI: Yes.

DR. STRAIN: It's very interesting.

DR. McCANN: One question, just to clarify a point. When you presented a group that you said had reduced cocaine use, how did you define that? How much of a reduction, what kind of a change?

DR. LAI: Thank you. This is a very important question. Basically, when the participant came in, we have to do urine test. If they're negative, we don't want them. We want urine positive. People would pay 10 bucks. So we said okay. One week later, they come over and have another test. If they are negative again, we pay 20. If next week they come back again, we pay 25.

Each time you are negative, we pay you more and more and more until you reach 80 bucks. However, if ever you show up with a positive urine test, the payment goes all the way back to 10 bucks.

So basically, the study participants, most of them are mathematicians. They know how to

Page 243

calculate and how to save money. This is crazy [indiscernible], all reduced. If the patient no show, we assume they are positive.

DR. McCANN: Reduced would be giving any negative urines?

DR. LAI: No. Here is the thing. We have 6 months, 120 days. Say we have 20 tests. We divide -- if we have one positive test, it's going to be 19 over 20. Every time you are positive, it's 20 over 20.

DR. McCANN: Okay. So were looking at a change over time.

DR. LAI: Yes.

DR. McCANN: You had a baseline period you captures, and you were just looking for a change --

DR. LAI: Yes, sir.

DR. McCANN: Okay.

MALE SPEAKER: Did you make any efforts to try to look for other potential causal variables of the outcome besides cocaine reduction, like diet? Could there have been other things that were going

Page 244

on in the subjects that might account for that outcome besides cocaine reduction?

DR. LAI: It's a very important question, but unfortunately we did not do it. We are extremely careful, just make sure we can do this study right because it's very sensitive. It involves money. You've got 80, right? The next time if you're positive, you're all the way -- you wouldn't get 80. You always would come back to get trained. Some time we have -- we just pay attention to the validity of a urine test. We will do later on if we have money.

DR. McCANN: Thank you very much. We've reached my 15-minute behind threshold, so now I'll introduce Ivan Montoya.

Presentation - Ivan Montoya

DR. MONTOYA: In the spirit of NIDA's interest, looking at the reductions of drug use associated with functional outcomes, we have some data sets that we wanted to start mining. So my presentation is just dissecting a little bit of data sets. I'm not going to present any final

Page 245

conclusions. It's just some thoughts about potential analyses and maybe ideas, to get some ideas from you, always to analyze this data and also maybe for you, too, that you use this data for future analysis.

The three questions that we have, one is, are quantitative urine cocaine results associated with psychosocial measures independent of treatment? Are changes in quantitative urine cocaine results associated with psychosocial changes, in this case, during treatment? And the third question, is a percent reduction of quantitative cocaine urine associated with psychosocial changes?

I'm going to go question by question, and this is what we have. This is the data that we have. We have data from 7 phase 2, double-blind, randomized clinical trials of different pharmacotherapies for cocaine dependence. They were conducted by NIDA through our contract program and ranked. And they all have quantitative urine toxicology analysis that were collected 3 times a

Page 246

week.

So those are the studies. The studies, some of them were for 8 weeks, and some of them were for 12 weeks. Those are the medications that were tested, and those are the sample sizes for each one of those studies. The studies are already published, and all of those studies were negative. This is the total sample size. It's about 1,353 subjects, and they also received manual-based, weekly, individual cognitive behavioral therapy.

For outcome variables, we have the ASI. We have all the domains of the ASI, and we also have the Clinical Global Impression, the observer, self-rated, severity an improvement, and the Brief Substance Craving Scale.

The first question, the idea was to look and see if we could see any signal by just doing an analysis, comparing the quantitative urine results and the different psychosocial measures. The idea was to see if lower cocaine use was associated with better psychosocial function independent of treatment. And for that, we used linear

Page 247

regression. So it's a pretty straightforward analysis.

Those are the results of the ASI. For the ASI, those are the different domains of the ASI. Basically, you can see this is the line that you need to look; this one. That is the linear regression with the confidence intervals. And as you can see, for most of the ASI domains, there were no significant difference. The only one that shows some trend was for the drug severity domain of the ASI.

The same thing for the Clinical Global Impression for the observed and self-rated. In this case, for the severity, there was observed a linear association between the quantitative urine drug use. I have to say that the quantitative urine drug, as was mentioned this morning, has a huge variability. And because of that, we have to convert it to a logarithmic 10, so all the data is logarithmic, not absolute values.

This is the Clinical Global Impression Severity. This is for the observed rate and the

Page 248

self-rated. For both, there was an increase in association between the urine test and the Clinical Global Impression. For the change, which is the other part of the CGI, there were no significant differences.

The other measure that we used was the Brief Substance Craving Scale, and for that one also, there was a trend. You can see with a greater amount of cocaine use or the greater the concentration of cocaine in urine, the worse the outcome, more craving.

With this, we went for the second question. The second question was, are there changes in quantitative urine cocaine results associated with psychosocial changes during treatment? That means, so those are the different trajectories of treatment that we can have in patients. We have treatment weeks here. We have cocaine use. Patients can decrease, they can be stable, or they can increase their drug use. The same is for the psychosocial measures. They can decrease, they can remain the same, or they can increase.

Page 249

In this case, for this analysis, what we did was for each participant during treatment, we calculated the slope of the log-10 benzoylegonine, and we compared that with the slope of the ASI domains, the CGIS course, the self-reported cocaine use, and treatment of retention. And for that, we calculated a regression coefficient and a p-value.

Here is what we have. What we have is the slope for, in this case, the log BE in each one of the ASI domains. You can see here for the ASI, for most of the domains, there's nothing except for the severity. We can see there are not [indiscernible] associations. This means that during time, during treatment, the patients who decreased drug use during treatment, their also ASI drug score decreases, but not the scores of the ASI.

For Clinical Global Impression, the same analogy. We have for both the slopes observed and the self-rated severity, not for the changes but only just the severity, there was a significant association. So the higher the concentration -- sorry. The higher the slope, the

Page 250

more changes in the treatment. And the reduction in log BE or concentration of drug use, the better the outcome based on the Clinical Global Impression.

This is for the Brief Substance Cravings Scale. It was marginally significant. This is the analysis looking at the urine quantitative benzoylegonine versus self-reported cocaine use. As you can see, there was a very good correlation between those two measures and prompted us to ask the question, what is the kappa coefficient, to the agreement coefficient, between the urine BE and self-reported drug use?

So this is the data looking at the kappa. The kappa, as you can see here, is .5. Between .4 and .75 is a good correlation. So I think there's a very similar agreement and what Kathy found in her study. Also, the result here looking at the self-report versus quantitative, new or used, those are the agreements, in general, pretty decent agreements. I know that clinically for me, an agreement of 50 percent in the patient means that

Page 251

50/50 presents chance of getting the right answer from the patient. So clinically may not be too terribly meaningful, but it statistically is very significant.

The other analysis that we wanted to look at the data was retention. We discussed during the day the retention is very important, and we wanted to see if the changes in the retention, if any of the changes were associated with survival and treatment. So here we have days of randomization. We have a year survival.

The data was divided in 4 groups by quartiles. The first group, the black group, is this group that had a very fast reduction in drug use. The blue group is a group that had an increase in drug use. And then the red group was a group that had sort of like a decrease in drug use, and the green group was the group that had changes increase and decrease in drug use.

The retention was shorter for those who decreased use more rapidly -- that's the first quartile -- and those who increased the use.

Page 252

That's the fourth quartile. That for me clinically is pretty interesting because those are the subjects that -- one group is the group that comes to treatment, do very well, and then they leave treatment. The other group is the group that do very bad, and they just leave treatment because they don't feel like they want to stay. That I thought was interesting to mention.

The third question is -- well, that question has been lingering around for a long time; what is the percent of reduction in drug use associated with psychosocial improvement? In this case, we look at the quantitative urine results, and we classify subjects as success, those subjects who reduced their drug use by 50 percent or more in the urine BE between baseline and the end of treatment. And failure, those that reduction was less than 50 percent.

Why 50 percent? Fifty percent was arbitrary. We did analysis looking at 25 percent, 75 percent, but 50 percent was the one that showed the best results. So we did a separate analysis

Page 253

for 8 weeks of treatment or 12 weeks of treatment because there were significant differences at baseline at the end of treatment in between those two treatment groups.

We looked at the end of treatment analysis, comparing weeks of treatment -- comparing the beginning of the treatment versus the end of the treatment. So we just compared the two points during the treatment. And subjects with missing data were excluded from the analysis because that was an exploratory analysis with multiple subjects, and we didn't include those that had missing data. And we did a step-down Bonferroni correction because of the multiple analysis.

The first one is the differences in mean ASI scores between treatment failure and success in that group that received 8 weeks of treatment. So I'm going to start with the 8 weeks of treatment. This is the ASI alcohol score and the ASI drug score. There was good news, there were reductions in the ASI scores. When we compare success and failure, there were no significant differences.

Page 254

This is for the rest of the domains in the ASI. There were no significant differences.

Now, for the Brief Substance Cravings Scale, when we compared the failure versus the success, we can see here that there is a significant difference. The delta, the reduction in urine BE and the delta here is significantly different. The same thing was for the CGIs. The CGI for both the observed rate and the self-rated there was a significant difference for the 8 weeks of treatment.

Now moving to the 12 weeks of treatment, this is the ASI. The only significant difference was for the drug score, which is the only significant difference that we found in the ASI. I don't consider that very meaningful. But for the other scores, the craving scale and the CGI, they all were significant. There were very significant differences between success and failure, as I said, defined by 50 percent reduction in drug use.

In summary, what we have is the question about quantitative urine results independent of

Page 255

treatment, in the ASI, we didn't find anything. The CGI, there may be some association. For the second question, any changes during treatment, for the ASI, not very much; for the CGI Severity, there's possibly a [indiscernible] association. Self-reported cocaine use, definitely. And for retention, there's also a nice [indiscernible] association.

Finally, for the percent reduction in cocaine, quantitative success versus failure, for the 8 weeks and the 12 weeks, what we found mostly was the differences in craving and the CGI. So in conclusion, the ASI may not be a sensitive measure to treatment change, at least that's measured by urine quantitative BE. The CGI appears sensitive, measures that may be associated with treatment success. Self-reported reviews is somewhat reliable, as I said. And cocaine use affects retention and treatment, the slope of the cocaine use. The 50 percent in cocaine in urine quantitative with treatment success is associated with improvement in the CGI and craving scales in

Page 256

8- and 12-weeks treatment regimens.

I just want to finish with an advertisement to advertise two funding opportunity announcements that we have open right now. One is for competitive revisions, what used to be called competitive supplements. And I know that some people in the audience have already submitted applications to this funding opportunity announcement. The second, which was recently published, is a program announcement for the R21 or 33, looking at reductions in illicit drug use and functional outcomes. Thank you.

(Applause.)

DR. McCANN: I'll take the liberty of starting with the first question. And I know better than to ask you what else have you done, have you done this or that. I know you've presented what you and Shou-Hua have pulled together. What I have in mind might be an idea for something additional we could look at.

Where you see the improvement in CGI and craving scales, in people have reduced by

Page 257

50 percent or more, certainly a group of the people who've reduced by 15 percent or more will have quit. So I'm wondering how much of that would have been driven by people within the people who reduced 15 percent or more who actually quit and how much would still be apparent in people who reduced but didn't quit.

In a number of the slides that Dan and Raye showed, they'd show people with heavy drinking, with moderate drinking, versus abstinent. I guess one of the reasons I'm asking is as challenging it may be to achieve abstinence, it's a heck of a lot easier to work with in some of our trials and really capturing percent reduction in use.

DR. MONTOYA: Yes. I think the number of urine samples that are negative are pretty small, but still I think it's worth doing that analysis. It's probably good to have a third group with abstinence. As I said at the beginning of the presentation, the data is so rich that there might be many opportunities for doing many types of analyses, including not only abstinence but also

Page 258

looking -- like Kathy and Brian presented, like the last 3 weeks of treatment, if that supports the same results. There are many other types of analyses, looking at different percentages.

FEMALE SPEAKER: I have a quick question. I'm trying to understand the 50 percent reduction. So the baseline was a urine sample taken at study entry, and then there was a subsequent urine sample. My question, there would be like a lot of sources of variation, how recently they last used, how pure the cocaine was, and so on and so forth.

How comfortable are you that a 50 percent reduction of quantitative benzoylegonine between sample A and sample B is actually a 50 percent reduction in their overall cocaine use?

DR. MONTOYA: I am not comfortable at all. It is just to measure -- there may be many clinical factors associated, the quality of the drug use. The variability, as I said, is huge. There are lots of limitations. I think the data is this, and it needs to be further mined.

FEMALE SPEAKER: Did you have some other

Page 259

measures of reduction based on the timeline follow-back or other kinds of things that you explored?

DR. MONTOYA: Yes. We have self-reported drug use collected by timeline follow-back. And that's how the self-reported drug-use data was presented. And as I said, correlated -- or at least in agreement with the BE was acceptable. It wasn't too bad.

MALE SPEAKER: You had a graph with the survival curves. There are four groups. And I think you were making some -- I didn't quit catch a conclusion. There's something about this is an implication for what happens when people drop out. Are they dropping out because they are getting better as opposed to getting worse? Did you say something to that?

DR. MONTOYA: The conclusion of this slide is that people are dropping out for two reasons: because they get better or they get really bad.

MALE SPEAKER: Could you tell like the relative prevalence of each of those reasons of the

Page 260

groups? I don't know. In alcohol, we kind of assume when people drop out, they kind of are getting back. I don't know if it's different here.

DR. MONTOYA: I'm sure they have the numbers, but I don't remember.

MALE SPEAKER: Each one is a quartile, right?

DR. MONTOYA: But the quartile is a quartile of the division of the data by log B -- by the slopes. And this the quartile of number of subjects, the quartile of the reduction, or the changes.

MALE SPEAKER: Oh, okay.

MALE SPEAKER: So there could be very few people --

DR. MONTOYA: It's the slope.

MALE SPEAKER: So there could be very few people in the people that drop out because they're getting better.

DR. MONTOYA: I think the sample size is big enough -- Shou-Hua, do you remember?

DR. LI: Eleven hundred.

Page 261

DR. MONTOYA: Yes, I know. But the sample size in each one of the quartiles --
DR. LI: I think each quartile is one-quarter.
DR. MONTOYA: We have the data. I can look at the data.
FEMALE SPEAKER: We always see a lag in terms of improvement in social functioning. It takes a while I guess to get your job back in line and getting along better with your family and so forth. It doesn't happen right --
DR. MONTOYA: Yes. I think that's a very good point. Perhaps 8 or 12 weeks of treatment is not enough to see changes in the ASI, specifically for legal, some of those changes. In fact, some changes in the ASI are desirable. I was talking with Phil the other day, and in one of the cases, there was an increase in the medical domain, which means that the subjects are now more aware -- the patients are more aware of their medical problems and they are seeking treatment.
That's why I'm so cautious about this data

Page 262

and the interpretation of this data, just presenting the results. But the interpretation can go in many different directions.
DR. STRAIN: Should we move to the discussant?
DR. McCANN: We'll introduce Phil Skolnick here to give us a discussant rant. I saw Lewis Black recently, so you have a lot to live up to, Phil.
Discussant - Phil Skolnick
DR. SKOLNICK: I didn't bring any slides, but I have a couple of remarks. It would be hard for me not to make remarks at this meeting.
We're here to think about the clinical benefit of -- how to evaluate clinical benefit in stimulant trials. I think the operant, the word that's missing, is a clinically meaningful benefit. The tendency in the past has been to focus on psychosocial outcomes, and that's valid, and you could have a clinically meaningful benefit from that.
The issue, though, I think from Ashley

Page 263

Slagle's talk, is to try to move towards more proximal measures that can be associated with abstinence or reductions in use. And I think Dr. Lai's talk, it's really impressive that you can relate something over time with a change potentially in atherosclerosis, something which has a medical benefit. You can tell a patient, if you take this drug, and you in fact over time reduce your cocaine use or you eliminate your cocaine use and have a reduction in the risk of atherosclerosis. I mean, that's a really significant outcome. So we're still in early days.
One of the issues I think that Dr. Lai's data brings out is the term of our clinical trials. Currently, most of the trials are 12 weeks or 16 weeks, and that really may be too short of time to capture some of the changes that we want to see, either in a medical outcome or by a psychosocial outcome. And I think the more distal we get, the more difficult it is to really evaluate that in a time frame of a 12-week trial or 16-week trial.
There are ways to modify the trials to make

Page 264

them longer, but then you have the issues of if patients are actually responding to the treatment, they may really want to get a life and a job and not show up twice a week to give urine samples and things like that.
There's one other thing I wanted to say; well, there are a couple more things. But one of the things I wanted to say was we talked a little bit about stratification. One of the things that's interesting, when I listen to some of the data, was that you have data from trials that are non-medication trials. And you have some interesting data that Kathy presented and Brian presented, which are a mix of therapy trials and pharmacotherapy trials. And even though we have some behavioral interventions in our pharmacotherapy trials, you wonder if in fact the outcomes are identical, if some of the data from the therapy trials alone would really obtain for a pharmacotherapy trial. We don't know that, and we really have to think about that.
The other thing I wanted to say just in

Page 265

concluding -- two things. Sorry. The first one is Ivan mentioned that we have an FOA out for outcome measures, basically reduced use. And I would encourage those of you in the audience that do that kind of research to think about applying for that grant. It's a very high priority for us at NIDA.

The final thing, it's more of an aspirational statement, is that part of the issue that was brought up earlier today is that we don't have any medications that are effective, or don't appear to be effective, for cocaine or methamphetamine use disorder. But one thing that has really sort of been ignored over the years is stratification based on compliance, meaning medication compliance during the trial.

From what we've seen in the trials where we've actually tried to measure compliance or a snapshot of compliance, in the substance-use disorder trials, it's such a low level of compliance and so little agreement with traditional measures, which is pill counts or self-report, that we may have had successful medications and that

Page 266

signal was masked because people took the drug and threw it in the toilet instead of taking it.

So one of the things we can do, especially in phase 2, where it's not a pivotal trial, is just stratify based on adherence in both the placebo group and the medication group, and see in fact if there's a signal. And that might down the road help us enrich our populations perhaps, or at least it's a starting point to talk to the regulators about how to best work these medications through.

That's it. Really a good session. It's been a great meeting. Thank you.

DR. STRAIN: Thank you.

I wonder if we might take -- we're close to a break time, I know, but maybe take five minutes for questions for any of the presenters, including the discussant, because I think there has been a lot of really good material here.

DR. McCANN: And especially for Brian because I did cut him off without allowing any questions when he got done speaking.

DR. STRAIN: Feeling guilty about that,

Page 267

David?

DR. McCANN: Yes, I am.

(Laughter.)

Q&A - Group Discussion

DR. STRAIN: So questions for any of the speakers, including Brian.

MALE SPEAKER: I guess I had a question for Brian. It was a great presentation, and I liked how you compared all the different outcomes to determine which ones might be more sensitive. I guess the issue of how to handle missing data is such an important one, especially when you're comparing across dichotomous versus continuous outcomes. And if there's differential dropout in the treatment arms, that can really impact the effect of the imputation that's done.

For instance, if each missing day -- each time you're missing an outcome in repeated measures, that can really -- if you're looking at a continuous outcome measure, if there's differential dropout, it can really punish one arm, versus like a dichotomous outcome, it's not as bad if you have

Page 268

missing data, assuming this really conservative imputation scheme of missing equal failure. That's what we find in our alcohol trials.

I guess my question is, did you guys try to handle missing data in any way in yours yet? I know there's many different ways to skinning a cat, but --

DR. KILUK: The paper that came out on drug and alcohol dependence, we operationalized each of the indicators and said how we were handling missing data. As Kathy talked, there were multiple ways. So we didn't look at it every different way that you could have calculated it, although we did come up with these reasons for why we chose the way we did or for handling other missing data, where if we're looking at end of treatment, is it the last 4 weeks they were in the treatment period or is it the last 4 weeks of when the treatment period is supposed to be, whether they were there or not.

In lots of ours, we just looked at when the people were actually in the treatment and tried not to sort of negatively impact those that dropped out

Page 269

if there was differential dropout because that is a big factor as well. I don't know if that answered your question. I mean, there are multiple ways to look at missing data. We operationalized it one specific way, and that's how that came out in ours.

MALE SPEAKER: I guess the first question would be to figure out is there differential dropout. If yes, then you really have to be real careful and try to apply -- I don't know. Sensitivity analyses are -- because there's no good way to handle missing data. There's no magic bullet for it, so you have to run the analyses using different imputations and see if the treatment effect size matters.

DR. CARROLL: And that's essentially what we do. And I think it was only really in that one trial where we got differential attrition by treatment condition. So that helps, and then -- this crazy thing that we do is chase people down and try to get them, whether or not they dropped out. We have sizeable numbers of those randomized, non-starters and dropouts.

Page 270

And it is. It's really interesting. I'm not sure that -- in most cases when people leave, they're not doing well, but not always. It's sort of a range that happens when people drop out, and it makes a lot of difference about whether you include the data post-dropout or not. It really does. And it's not always in the ways you would expect.

We obsess about it, and all you can do is really try to minimize it, and then really just chase those folks down. And it does hurt some more than others. In jumping into this, it would be best if we were doing it all roughly the same way, or at least having a consistent -- making the assumptions transparent I think is really the only way to do it.

That's why I think I'm so passionate about just come up with something that we all calculate the same way because right now it really is apples and oranges, behavioral trials and pharma trials. We often can't compare outcome. One of the things I'd like to come out of this meeting is that we

Page 271

just agree to start somewhere; that we all do it more the same than we have been in the past.

DR. STRAIN: Perhaps on that optimistic note, we could take our break.

Valorie, are we next-door? Why don't we plan to reconvene here at 3:15 as scheduled, and we'll be doing our last session. Thanks to all the presenters.

(Applause.)

(Whereupon, a recess was taken.)

DR. STRAIN: Why don't we go ahead. Rachel Skeet's going to now talk to us. And I don't have your title in front of me for your talk, but I'll let you introduce it. Thanks.

Presentation - Rachel Skeete

DR. SKEETE: There's the title. Good afternoon, everyone. Thank you for this opportunity to be a part of this meeting and share some lessons learned for trials in other addiction areas. I was the clinical reviewer for the new drug application for probuphine, and this is for maintenance treatment of opioid addiction.

Page 272

Today, I'll be discussing key lessons that our review team learned as we reviewed this application. I'd like to stress that this will not be a discussion of probuphine. Instead, I'll be discussing the lessons we learned during our review that may have implications for stimulant trials. And I'll be referring to the probuphine program only to illustrate these lessons. For this talk, I'll be using publicly available information only. And before proceeding, I'll make a disclaimer that these are my views and not necessarily those of the FDA.

This is an overview of the key points that I'll discuss. The lessons learned are listed here as discreet concepts, but you'll see that there's some overlap between the treatment design elements that I discuss. As I mentioned, I'm using probuphine as a case study. I'll be using the probuphine example to illustrate four main points that were challenges for us during our review of this application. As I mentioned, I'm using probuphine as a case study.

Page 273

For the four main points, the first has to do with the choice of treatment responder and failure definitions. In the case of the probuphine trials, the protocol-specified treatment failure definition was based on rescue medication needs. Drug use behavior was not considered in definition. But at the same time, a measure of drug-use behavior but not a measure of rescue medication needs was used as a protocol-specified criteria for evaluating treatment response or efficacy.

The second point is that trial design can make the placebo failure almost certain, and in so doing, it can bias a trial towards a positive result. The way this happened with probuphine is that both trial groups were allowed rescue medication based on withdrawal and craving symptoms, but the placebo group is almost certain to have higher rescue probuphine needs.

As I mentioned earlier, treatment failure was based on rescue medication needs, and so the placebo patients often met the thresholds for the treatment failure definition, and then they were

Page 274

discontinued from the study. So once a patient was discontinued, their urine samples from that point on were considered as positive. And you'll see that this was much more common in the placebo group.

Thirdly, treatment retention or a longer time in treatment on study does not always mean that patients are improving. Here you see that patients remain in a trial and presented for urine collection visits 3 times a week for 6 months, but they were still continuing to use illicit opioids.

The other point with time and treatment on study has to do with grace periods. We allow sponsors to include grace periods in trials where we ignore data for the first few months. This is because we think that patients may need some time early on in treatment to engage in treatment. However, we learned -- and you'll see -- that it's not guaranteed that patients will improve over time.

The final points or lessons that we learned have to do with the response profile approach.

Page 275

This was actually described earlier -- thank you to the people from NIAAA -- when they showed you their curves of cumulative proportions of responders. What we learned from the response profile is that showing a difference in the curves alone is not enough. We still need to understand the relationship between drug-use patterns and clinical benefit.

Finally, we learned that the way in which you display the results can really influence how the results are perceived, so the choice of how to display the findings bears considerable attention.

I'm not going to expand on each of the four lessons learned using the probuphine case study. I'm first going to give you a brief background on probuphine and the trials. Again, this is not a discussion of probuphine. I'm only providing this background information to give you some context for the discussion.

Probuphine is an implantable formulation of buprenorphine. It provides sustained release of buprenorphine for up to 6 months. I'll be

Page 276

referring to the individual implant units you see here as rods, and each of these rods contains 80 milligrams of buprenorphine. So treatment with probuphine involves initial treatment -- in other words, induction -- with a sublingual or transmucosal form of buprenorphine, and that's to reach a target dose of 12 to 16 milligrams per day for at least 3 days. Then 4 rods are inserted into the upper arm. And then based on supplemental buprenorphine or rescue needs, an additional 5th rod can be inserted. The rods are taken out in 6 months, and treatment can be continued by implanting into the opposite arm for another 6 months.

There were 2 probuphine efficacy and safety trials, which I'm describing for context only to illustrate the lessons learned. Remember, this is not a discussion of probuphine. PRO-805 and PRO-806, which I'll sometimes call study 5 and 6 for short, were 24-week, phase 3, randomized, double-blind, placebo-controlled trials in opioid-dependent patients. Study 6 had an

Page 277

open-label sublingual arm, however, I'm limiting the discussion of the trials to the probuphine and placebo arms and to a discussion of efficacy only.

Like the proposed dosing procedures I just described, in these trials, subjects were initially treated with sublingual buprenorphine to reach a target dose of 12 to 16 milligrams per day. Probuphine or placebo rods were then inserted for 24 weeks. An additional rod was added if protocol-specified, supplemental sublingual buprenorphine needs were met.

Patients could receive supplemental or rescue buprenorphine for withdrawal symptoms when subjects scored more than 12 on the Clinical Opioid Withdrawal Scale or if they had what the study called cravings and endorsed more than 20 millimeters on the Craving Visual Analog Scale, or if they requested buprenorphine, and the request was seen as appropriate by the investigator. Now, in study 5, only one criterion needed to be met, but for study 6, all three needed to be met.

Rescue buprenorphine was obtained at the

Page 278

clinic or pharmacy, and patients were required to get a dose increase or insertion of a 5th rod if they received supplemental buprenorphine on 3 or more days for 2 consecutive weeks or 8 or more days over 4 consecutive weeks.

During the trials, urine samples were collected 3 times a week and investigators were blinded to the urine toxicology results. Patients also provided self-report of illicit opioid use approximately every 2 weeks. Patients were discontinued early for treatment failure and non-compliance as well as other reasons.

We considered treatment failure and non-compliance as two of the key reasons that relate to efficacy. So subjects were considered treatment failures if after they got a dose increase of a 5th rod, they still met the same protocol-specified criteria, that I mentioned earlier for supplemental buprenorphine, that required them to get that 5th rod. They were considered non-compliant if they missed 9 urine visits in a row, 6 counseling sessions in a row, or

Page 279

by investigator decision.

The primary endpoint was a cumulative distribution function, or CDF, which is consistent with what we consider a response profile. It was a cumulative distribution function of the percentage of opioid-negative urines over the 24 weeks of treatment. For a urine sample to be considered opioid negative, both the urine sample and self-report around the time a particular urine sample was collected had to be negative. Missed samples were considered positive.

This slide shows how the cumulative distribution function works. I won't spend too much time on it because it was discussed earlier; thanks again. The CDF looks at opioid-negative results cumulatively. For example, in this histogram illustration, it's showing patients at each of these categories of opioid-negative urines in a treatment arm.

Where you see the 8 percent of patients who had 95 percent of their urine samples negative for opioids, when this is looked at in a CDF, these

Page 280

patients making up that 8 percent would be counted in the category of patients with 95 percent urines negative or more. And they would also be counted in the at least 85 percent category, at least 75 percent category, and so on, because if they have at least 95 percent of their samples negative, they also would have satisfied all the other lesser categories.

This slide provides an overview of the primary efficacy results that I'll be using to illustrate each of the lessons learned, from opioid addiction trials and this case study. The primary efficacy measure, again, was a CDF. It was the primary endpoint that we agreed upon. This type of analysis has advantages and involving clinical trial design areas, like the are of opioid addiction, where it's difficult to establish a definitive responder definition.

We were unable to come to an agreement about an appropriate responder definition in this case, so we saw the response profile as one way that we could avoid excluding those patterns of use that

Page 281

could represent a clinically meaningful change in drug-taking behavior, but may not be part of a fixed definition of response.

When we consider the use of a response profile for the analysis, though, we hoped and we envisioned that the curves would separate at those points along the X-axis that represent abstinence or near abstinence. We're talking about this area right here. And that's particularly if we were allowing a grace period. However, as you'll see, the findings in the probuphine case were different than what we had expected.

These graphs, as well as additional analyses that I'll be showing you later on, were generated by David Petullo. He was the statistical reviewer for the probuphine application. On the left side of this slide is a graphical representation of the findings for each trial, and this is showing the CDF curves. On the right is a tabular summary.

Let's first look at the graphs. On the X-axis is a proportion of opioid-negative urine, and then on the Y-axis is the proportion of

Page 282

patients. The solid curve is a probuphine arm, while the dash curve is placebo. If you look at the 0.3 mark on the X-axis, this refers to 30 percent or more urine samples negative for opioids.

Looking at the proportion of subjects meeting this threshold, between about 40 and 50 percent of probuphine patients had 30 percent or more opioid-negative samples, while a little under 30 percent of placebo patients had 30 percent or more. Again, we had hoped to see a separation of curves on the right-hand side of the curve for both trials and higher proportions of patients achieving abstinence or near abstinence. However, what we saw instead was separation towards the left and middle of the X-axis, where these changes in drug-use behavior are less conclusive, particularly for study 5.

The tabular summary on the right shows the same findings. There were no abstinent patients and few near abstinent patients. You can see also that the placebo rates in study 6 are markedly

Page 283

lower compared to study 5, and this probably represents a higher dropout rate with a stricter criteria for receiving rescue medication.

Now on to a discussion of the specific lessons learned using these findings for illustration. The first lesson, again, deals with the treatment response and treatment failure definitions. This slide shows subject level urine toxicology data for day 5 -- for study 5. Each subject is represented as a point along the Y-axis. So when you follow a line across, you see all of the patients urine toxicology results over that 24-week period. A blue dot is a negative urine sample, red is positive, and a plus sign is missing.

As an example, the placebo patient, the first one on the very bottom that you see here -- and you might have to take my word for it -- had 1 opioid-negative urine sample, then 2 positive samples, and then was discontinued. From that point on, all the rest of the urine samples are considered positive. Ideally, on these graphs,

Page 284

you would see a lot of blue overall, especially on the probuphine side, but instead you see a lot of red for opioid positive urines.

Drug-use behavior, again, based on urine toxicology and self-report data, was used to evaluate efficacy, and that's what you're seeing here. But it was not considered in the protocol-specified treatment failure definition. We find it difficult to interpret these results because there a considerable number of patients who continued to use throughout the entire treatment period, even though they didn't meet protocol-specified treatment failure definitions.

We found the subject level urine toxicology to be similar for study 6. And here in the probuphine arm compared with study 5, there's arguably more evidence of opioid use. So when we reviewed the results of these trials, we were left with a number of uncertainties. We wandered how the investigators would have assessed treatment response if they had been aware of the urine toxicology results. Also, there were observer

Page 285

rated clinical global measures that were assessed as secondary outcomes, but the urine toxicology findings again weren't available to the investigators. So we questioned the utility of a global assessment measure in this context.

Finally, when you look at the number of patients who continue to use illicit opioids during this study, these are those patients with red dots most of the way if not all the way through. We wondered how the results might have differed if drug-use behavior was part of the treatment-failure definition.

Another lesson we learned relates again to the potential for a trial to almost guarantee placebo failure, and when this happens, the trial can be biased toward a positive result. To illustrate this plan, I'm using subject level analyses of rescue buprenorphine use during the trials.

This analysis is pretty similar to the ones you just saw for the urine toxicology results. Here, though, subjects are aligned in the order of

Page 286

when they discontinued from the trial. X's represent the point where a patient discontinued. Red dots are for days when a patient use rescue buprenorphine, and dashes are for days with not use. When you look at the placebo patients, you can see that many discontinue early, and you can also see a lot of red prior to the early -- the continuations here singling that there was a lot of rescue buprenorphine use up to that point.

Remember that patients were discontinued when they met the protocol-specified treatment failure definition, and treatment failure was based on use of rescue buprenorphine, but it wasn't based on drug-use behavior.

Once these patients discontinued from the trial, they weren't providing additional urine samples, so those samples from then on were counted as missing. Missing urines were considered as positive, so this means that these patients were judged to have opioid-positive urines from the time they discontinued all the way to the end of the 168-day or 24-week window.

Page 287

We found this imbalance in placebo failure and dropout difficult to interpret. This was because it appeared to us that the imbalance stemmed from opioid withdrawals symptoms for those on placebo and also from the treatment failure definitions. It seemed that it had less to do with an effect of drug-use behavior, although drug-use behavior was used to evaluate efficacy.

The reason we thought this is because, as I mentioned earlier, patients had to reach a target dose of 12 to 16 milligrams per day of sublingual buprenorphine for 3 days before they received probuphine or placebo rods. After the placebo subjects reached its target dose and then were abruptly discontinued, you can see how they could need a lot of rescue early on. But they ended up meeting the protocol-specified treatment failure definition by virtue of these rescue needs only.

On the probuphine side, patients stayed in the study much longer, but we saw before that many continued to use opioids throughout the entire treatment period. Our understanding here was that

Page 288

they were receiving buprenorphine and didn't need as much rescue, so they did not meet the protocol-specified treatment failure definition. On the other hand, based on evidence of illicit opioid use that we saw earlier, many of the probuphine patients didn't appear to be treatment responders to us.

In study 6, there appear to be even more placebo dropouts than there were for study 5 and less concentration of rescue buprenorphine use among the early discontinuations relative to study 5. We think this may be the case because of the more restrictive rescue buprenorphine criteria in study 6. Again, all three criteria were needed for study 6, and one criterion was needed for study 5.

The next lesson we learned during our review was that treatment retention does not always equal improvement. This is the subject disposition summary for each of the trials. It summarizes the reasons for early withdrawal that we consider to be related to efficacy. In both studies, twice as many subjects in the placebo arm withdrew early

Page 289

compared to the probuphine arm. Treatment failure seems to be an important driver here with non-compliance also contributing to early withdrawals.

Neither the protocol-specified definition for treatment failure nor for compliance considered drug use. From the last set of graphs, we saw that patients continued to use opioids throughout the entire treatment period and were not considered treatment failures. Therefore, even though patients in the probuphine arm had longer time on study, they still didn't necessarily improve over time.

Another view of ours related to time on study --

MALE SPEAKER: Quick question.

DR. SKEETE: Yes?

MALE SPEAKER: What do you mean by non-compliance? We usually think about not taking your meds is non-compliant.

DR. SKEETE: Right. I should say protocol-specified definition of non-compliance.

Page 290

So again, that was -- so if they missed 9 consecutive urine visits or they missed 6 consecutive counseling sessions.

MALE SPEAKER: Thanks.

DR. SKEETE: This now talks about looking at grace period. Another thing that we look at with time on study is that patients may need some time to engage in treatment. We believe that if we allowed for a grace period where we ignored the first few months in the analyses, it would help to show a treatment effect. However, I review of the probuphine application taught us that patients are not guaranteed to improve over time.

The two graphs on this slide show you cumulative distribution function of the percent of opioid-negative urines when all the data are included -- that's the top graph -- and when we allowed for a 4-month grace period. That's the bottom set of curves. For study 5, whether or not we allow for a grace period, the overall efficacy findings are similar.

For study 6, we also see that a similar

Page 291

pattern of drug-taking behavior is evident, regardless of whether the grace period is allowed. So we saw that it's not inevitable that patients will improve substantially over the course of the treatment period, and the grace period won't necessarily help in demonstrating a treatment effect.

The last set of lessons we learned have to do with the response profile approach. Remember that the response profile is essentially a way to look at an entire continuum of treatment responses over a range of responses. We're interested in individual treatment responses in these addiction trials, so the response profile can be advantageous in showing individual responses while not limiting treatment response to a single responder definition. It allows some flexibility in defining clinically meaningful changes in drug-taking behavior when a specific responder definition is difficult to establish.

These are the main set of curves for the primary endpoint. These results differ from what

Page 292

we had anticipated when we agreed upon this approach. We expected to see more compelling results and separation of the curves at the right-hand side of the X-axis among a larger number of patients. For both of these studies, there was a statistically significant difference in favor of probuphine. The difference was in the more minor changes in drug-taking behavior for study 5.

We didn't consider these overall changes in drug-taking behavior to automatically translate into clinical benefit, so we learned that when using the response profile approach, demonstrating a difference in the curves is not enough. We still need to understand the relationship between the changes in drug-taking behavior, which is a surrogate endpoint and clinical benefit.

This is the last graph I'll be showing you today. I'd like you to take a quick look at this graph and see what you gather from the display. So what's your interpretation at first glance? When you take a quick look, you may think that there were a number of patients achieving abstinence or

Page 293

near abstinence, especially when you look at the Y-axis, and you see the higher percentages of patients, for example, those 90 percent or above. And you see that plotted against the higher proportion of opioid-negative urine samples along the X-axis.

It may even look as though the placebo arm, which is in pink, is doing a little better, especially if I told you there was a statistically significant difference between the two curves. However, this graph is just another representation of all the information that you've seen so far, so why does this one tell a different story from the others?

This is the original graph that was used to demonstrate the sponsor's results. This display resulted from graphing the proportion of subjects who submitted a certain level of negative urine samples or fewer. All the other curves you've seen have graphed their proportion of subjects who submitted a certain level of negative urine samples or greater.

Page 294

On this graph, approximately 50 percent of subjects in the probuphine arm had no more than 30 percent urine samples negative, while approximately 70 percent of placebo patients had no more than 30 percent urine samples negative. This shows you just how much the choice we make in presenting the graphs impacts our understanding of the results.

In summary, the key lessons we learned were that treatment failure and responder definitions are difficult to interpret when they don't take into account all the available information in the study that is considered important. For placebo failure, trials can be designed in such a way that they all but guarantee placebo failure, and that could bias the trial toward a positive result.

It's difficult to interpret a treatment effect in this context, particularly if the treatment effect is also equivocal in the treatment arm. For treatment retention, we learned that remaining in a study or staying on treatment longer does not automatically amount to clinical benefit

Page 295

or improvement. At the same time, it doesn't appear that patients will necessarily improve over time, so allowing a grace period for patients to engage in treatment will not necessarily help to demonstrate a treatment effect.

Finally, with the response profile approach, from examining the curves, we've learned that any difference does not equal a clinically important one. We need to understand the relationship of the findings to clinical benefit. We also saw that because of the impact the display of the results has on your perception of results, the choice of parameters used for the graphical display bears close attention.

This ends my presentation, and I'd like to thank you for your attention, especially at the end of the day. And I'd also like to acknowledge those in DAAAP who provided leadership and guidance both for the review and for this talk, and also for the many reviews both in CDER and CDRH, that's the drug center and the devices center who participated in this drug device review, and our statistical

Page 296

reviewer. And if there's time, I'll open for questions.

(Applause.)

Q&A - Group Discussion

DR. STRAIN: Thank you, Dr. Skeete.

This is open for questions, this talk. George?

DR. WOODY: George Woody. One of the real advantages for alcohol, you know how much alcohol they're taking. But when they're using heroin, you don't know how much heroin they're taking. You could ask them how many times a day they're using or how many days they're using, but more typical, how many days of use I think is the most common thing.

Within that, theoretically at least, you could show use, but there could be a reduction in the amount of use because the urine test will register. Sometimes I think that with -- like naltrexone -- or certainly with methadone, you see a lot of continued use in methadone. But it's almost like you take an alcoholic and turn him into

Page 297

a social drinker with some of the things. But when you're evaluating the opioid studies, it's a little hard to get that because you don't know how much, just as a comment. I don't know how to get around that.

DR. SKEETE: Right. If you have suggestions, we're always open to hear. That's one of the difficulties in looking at specifically opioid use.

DR. STRAIN: Kyle?

DR. KAMPMAN: You said that just because there was a difference in the use, that there wasn't necessarily a clinical benefit. But I kind of remember from that trial, weren't there a whole bunch of secondary outcomes that favored probuphine?

DR. SKEETE: The primary outcome -- I have to, again, remember this is publicly available information only, so I'm trying to remember all that was said --

DR. KAMPMAN: Maybe I shut my mouth.

DR. SKEETE: -- all the other things. There

Page 298

were other secondary outcomes that were evaluated and that did show some potential. But the way that the study was designed it was to look at that CDF, and it was powered to look at that.

Some of the secondary outcomes, including the one that I did mention, was also the observer rated clinical global impression. And that was hard to interpret because the investigators, when they were making those conclusions, weren't aware of any of the urine toxicology results. So we're still not sure how that would have been impacted or not had they been able to have that information in front of them.

Treatment retention was actually a secondary endpoint. So, yes, that was a positive outcome, but you can see the nuances of that is that they still were continuing to use over the time.

MALE SPEAKER: My understanding of the cumulative responder curves, there was a big treatment effect at certain levels of reduction, but you're saying that that's because there's differential dropout in the way of imputing the

Page 299

missing data in the placebo group really punishes -- the placebo group makes them look a lot worse.

What was the reason why we think that there is more differential dropout with placebo group? Do you have any suspicions?

DR. SKEETE: There are a couple of things. One is the way that the -- when they started treatment -- so both groups start on sublingual buprenorphine, and they have to reach a target dose because they need something before they get the implant put in. They have to reach a target dose of 12 to 16 milligrams. So they are stabilized on that dose or "stabilized" reaching that target dose between 10 to 16 days.

Once they get that target dose, and they're doing okay on that dose, they then get switched to the placebo or the probuphine rod. Now, the placebo group is essentially abruptly discontinued from treatment, and then the only thing that they get is the rescue buprenorphine, whereas the other group has a low level of buprenorphine that can

Page 300

manage withdrawal. So now there is sort of differences in that, and that's what we think is some of the issue there.

MALE SPEAKER: But there's been a way to do it -- I know -- a way to do it better?

DR. SKEETE: Well, if you have some suggestions -- that's one of the things we wanted to bring out is we wanted to be able to compare -- we wanted to be able to see if there was some way to look at it compared to placebo. And the rescue medication was actually put it there because it would have been considered unethical to just put people on -- get them up to a probably 16 dose, and then just abruptly discontinue, and then give them nothing.

We're open to hearing if there are thoughts on ways that you could see a way of looking at placebo or treatment effect while you can still now have this differential placebo dropout that we saw.

DR. STRAIN: David?

DR. McCANN: For the rescue medication, were they allowed to take that home at all, or did they

Page 301

have to take it on site?
DR. SKEETE: They had to take it on site.
DR. SKOLNICK: Phil Skolnick. Could you comment -- if you're allowed to -- on the difference between the phase 2 design and the phase 3 design? The sponsor must have been incented to go to phase 3 based on the end of phase 2 meeting. And I'm just wondering --
DR. SKEETE: How that happened?
DR. SKOLNICK: Yes.
DR. SKEETE: Some of that stuff I think is not publicly available.
DR. SKOLNICK: Well, if you can't, you can't.
DR. SKEETE: Yes. Actually -- yes. I'm not sure I can mention that one.
DR. STRAIN: Kenzie?
DR. PRESTON: For that particular trial, could they have chose retention in treatment as a primary outcome measure? Because really what, it seems to me, that you're comparing is a detoxification versus maintenance. And there's

Page 302

already literature to say that people do better with longer term treatment maintenance than detox.
DR. SKEETE: In other words -- so instead of using the response profile based on urine toxicology --
DR. PRESTON: Drug use.
DR. SKEETE: -- drug use, because they have used treatment retention instead. The only thing, though, is that we still have the concerns about what the treatment retention was saying. When we looked at the treatment retention results, we saw that they were staying in -- the people on the probuphine arm were staying in a trial twice as long as the other -- twice as often as the placebo group. But at the same time, they were using illicit opioids.
So it was hard for us to say that the treatment and retention alone was enough because even if they're staying in the treatment, they're still using illicit opioids. I guess it would -- you would need to demonstrate that treatment retention alone actually will translate

Page 303

to the clinical benefit. Because what we're seeing is that the use behavior is being used as a surrogate that we sort of understand to translate to clinical benefit. But the treatment retention alone, I guess we would have to see that it demonstrates that it's clinical benefit.
FEMALE SPEAKER: I guess I would like to put into the conversation the point of the lesson, which is the study design sounded good initially, but in seeing the results, we learned about the challenge. It wasn't so much a differential dropout rate. We had differential drop out due to the placebo and active treatments in a number of different therapeutic areas. In pain, for instance, it's classic. You see more dropouts in placebo due to lack of efficacy, adverse events and active treatment groups, really has an impact on how you impute.
Here, it was a protocol defined discontinuation and attention to whatever actual treatment assignment related dropping out may have occurred. And that's where it became challenging

Page 304

because unsuccessful treatment was defined based on the use of rescue, but not on the ongoing or continuous use of illicit drug. And it created an opportunity to differentiate the two treatment groups when neither was particularly successful.
So I think the point here is -- the lesson is when you define a treatment failure, A, what do you do with that person? Do you continue them in the study and collect information; and B, do you take all the proper endpoints into consideration in defining your treatment failure?
So I think that those are sort of -- the point being, we have to use this as an opportunity to understand the impact of choices. You don't want to create a study that fails to distinguish treatment effect if, for instance, there's a large amount of rescue. Right? Because that can mask a treatment effect. On the other hand, you don't want to exaggerate an effect by failing to take into account what might, in retrospect now, be useful to consider as part of your treatment failure definition.

Page 305

So if you're going to penalize, in a sense, the placebo group for needing rescue -- now granted, you would technically penalize both groups if they need a rescue. You should probably also -- I don't mean penalize, but I mean take into account. You should probably also take into account other things such as a certain amount of illicit drug use.

Those are the sorts of things that may be useful in thinking about design of the stimulant trials once there is a product. And the other issues that have been discussed in terms of entry criteria and everything else, if the paradigm chosen is to, for instance, use a treatment failure design, not to fall prey to a similar situation.

DR. STRAIN: Other thoughts?

(No response.)

DR. STRAIN: It's complicated. This is the -- my thought is it's easier to figure out how to not design a study -- and this isn't unique to this than it is to how to design a study. And that's what I struggle with, with this.

Page 306

Were you going to respond or say something further? Because I was going to go off to another question. It sort of gets back to this.

In part, this is an example of the design of the trial, but we're also focused upon the outcome measures. And I wonder for stimulant trials if there are specific lessons from this trial regarding outcome measures that could help inform us for stimulant trials, not just in terms of study design but in study measures.

DR. SKEETE: I think -- well, the study design will impact the measures. But one of the things is if you -- you might -- we've been talking today about whether abstinence is the only way we can go or are there other use patterns that might make a difference.

One thing that you may be able to take away is thinking about this response profile approach. If you were to go and look at stimulant trial and sent that way, what could you take from here, from the opioid addiction trial setting to say how you would apply that to a stimulant trial might be one.

Page 307

Another one, someone mentioned whether you could use treatment retention, but again, it still will go back to the study design, but again, to think about is treatment retention; how would that apply in a stimulant trial, for example, and what do you need to think about in terms of the trial design to be able to do that.

DR. STRAIN: I'm not sure if I can articulate this, but it strikes me that -- what we're trying to show is clinical benefit. So often what we're basing that upon is drug use, and what we're basing drug use upon is urine results. I guess from this study the difficulties are the compounds of supplement probuphine use, early dropout, and illicit drug use.

I think those are the three sort of things that are potentially impacting the interpretation of drug use, which is what we're saying is the measure of clinical benefit. Have I got that sequence right?

DR. SKEETE: Yes.

DR. STRAIN: One of the things I guess I'm

Page 308

getting my head wrapped around is are there other measures of clinical benefit that may be of value beyond drug use, endothelial damage or something, potentially.

Ken, please.

DR. SILVERMAN: This is Ken Silverman. I don't have an answer to that question. This study is confusing. I'm not sure what -- except one thing it does have in common with other studies that have been talked about today is that missing data is a big problem. And in fact, I think it's probably the biggest threat to the quality and validity of the outcomes than anything, maybe more than what measure you happen to pick.

Extended-release naltrexone, the Depotrex trial that showed that extended-release naltrexone increased retention. And when you impute missing samples as positive, it also looked like it reduced opioid use. But of course when you impute missing samples as positive, it also reduced benzodiazepine, amphetamine, marijuana, cocaine, and everything else that they tested for. And that

Page 309

publication was pretty thoughtful that they included different types of missing data, and only the missing positive showed these effects between the two groups.

In fact, I think the same problem exists for the extended-release Vivitrol study that was used to approve Vivitrol for opioid dependence. Vivitrol also increased retention substantially, and it looks about the same amount that it improved urinalysis results, which I think were also based on missing samples positive.

MALE SPEAKER: Thought in that particular study -- I worked on the Vivitrol program -- I don't think that there was any forced criteria where patients were ejected from the study on one criteria.

MALE SPEAKER: No, no. I didn't mean to suggest that. I'm not even talking about the FDA specifically. But if you looked at the urinalysis results, there's a big difference. Vivitrol extended retention. People who got Vivitrol were in treatment longer than people who got placebo.

Page 310

And then when you look at your analysis, when you compare the two groups on your analysis results, if you impute missing samples as positive --

MALE SPEAKER: What we did in that particular -- one difference I think also with antagonist treatment, as opposed to agonist, you can't really use and get high while you're on an antagonist. So there is maybe a difference between being assured that you're on a blocker.

MALE SPEAKER: I've got --

MALE SPEAKER: Just to say, for the criteria that we use for that particular study was that in order to be successful you had to show up at the clinic. You had to give a urine, and the urine had to be negative. In any other situation, you are considered positive. So there was no kind of --

MALE SPEAKER: Well, still -- we shouldn't argue about this specific study, although I still think that that study imputed missing samples as positive.

FEMALE SPEAKER: I think this is a very important point, and perhaps you all are familiar

Page 311

with the National Academy of Sciences' report, from a few years ago, that looked at the issue of -- it started off as what's the right way to impute missing data in a long-term study? And the answer was, minimize missing data because all methods of imputation suffer from some type of problem.

MALE SPEAKER: That's right.

FEMALE SPEAKER: And there is additional layers there. So when imputation is necessary rather than a single imputation, for instance, last observation, baseline observation, positive, what is another approach? There's something called the multiple imputation method, and I refuse to tell you how I understand that because it's the kindergarten version. But that's one approach.

But more importantly is the way it's been stressed to minimize missing data, even if someone goes off treatment, to keep them in the study and collecting data from them. Now, I grant that this population in contrast to a pain population may be harder to do that with. But the idea of different incentives to bring them in, that's I guess the

Page 312

challenge for this therapeutic area, is how do you get these people in even if there's going to be a positive urine that they're trying to hide?

Those are the kinds of things, if people have ways where there have been better successes there, that would be a best practice.

MALE SPEAKER: Yes, that's my point. The reason that I mentioned this is mostly to say that probably the most important thing that we an agree upon is that you have to -- you can't accept studies that have differential retention, and you have to have comparable and high rates of collection, and you have to find methods to do it.

I think if you looked at the results that Kathy Carroll and Brian presented, they have like these long-term follow-ups, 1, 3, 6, whatever. I think you've got over 80 percent of those urine samples for those. The problem is you can get samples like that -- and we do as well, for like monthly samples -- but when you try and do it for Monday, Wednesday, Friday urine samples, you cannot do it.

Page 313

So developing methods and criteria that says, okay, this is an acceptable trial if you get comparable retention, high rates of collection --

MALE SPEAKER: One other interesting thing that occurred in that --

DR. STRAIN: Hold on one second. Hold one second because I want to thank Dr. Skeete. I think we're moving into a more general discussion, so you can sit down if you'd like.

MALE SPEAKER: I just wanted to make just one other follow-up, which I think was a really interesting lesson learn that we had from the Vivitrol opioid program. We did specify for that trial a response profile. And at the end of the day, though, if you looked at the subset of patients who stayed in for the entire study and were completely abstinent on Vivitrol, there was a significant difference from placebo.

I think my interpretation was that was something that was important to the FDA reviewers. So although maybe we came at the trial from a different -- we were looking at response profile,

Page 314

and I think there was a preference on the FDA to look at that subset that were completely abstinent. There was a situation where I think we agreed to agree. So that may also have applicability when thinking about trials and stimulant.

DR. STRAIN: Thanks. Kathy, you had a comment. And you need to get near a microphone, and then Dan.

DR. CARROLL: I think we can think about these things a little bit differently because the mind-set really is, they're gone, they're gone, and it doesn't have to be that way. You can collect data -- you have to train the staff and you have to train the patients, too, that participation in treatment doesn't have to be linked to participation in the trial. And you can have them go to the labs. They don't have to come in to the nasty clinic and get yelled at by the staff to do it. You can interview them in different places. You can do the labs.

There are ways you can do this, and I'm not clear why we invest so much in these trials, and

Page 315

then don't go that little bit more to get the data, which just kills it. I think we have to not use trials that get differential or get less than 80 percent of the samples because you just don't know it.

DR. STRAIN: Dan and then David.

DR. FALK: I guess my question was -- I don't know if Rachel -- if the FDA has a position on this. But with responder profiles, that's a very kind of post hoc way -- you can't declare -- can someone declare up front for a phase 3 trial, like we're going to look at abstinence or we're going to look at all the other possible cut point and kind of decide.

So the general question is, how would the FDA -- could the FDA ever prove something for a medication based on a cut point that wasn't defined as a primary just from looking at the responder profiles? Or maybe even a more general question, how does the FDA use responder profiles?

FEMALE SPEAKER: So I'm going to speak not as the FDA. I'm going to discuss responder

Page 316

definitions, continuous responder functions, and the analysis. There are statistical methods that can be applied to evaluate the separation of the curves. So there is not a specification of a cut point, there is just an analysis of the separation of the curves.

I can't describe the details, I just know they exist because they are proposed in other therapeutic areas. The question, though, or the relevant point from the probuphine example is that that type of analysis, the separation of curves, may not actually be the right outcome for all clinical situations. If the value to the patient is dependent on the far right of the curve, then the analysis needs to reflect that. If any separation along the curve that meets the statistical endpoint is sufficient to confirm clinical meaningfulness for the endpoint reflected in that curve, then that is also appropriate, but it's going to change by clinical setting.

So responder definitions and responder data can be analyzed using this continuous function. It

Page 317

can be changed into a dichotomous outcome. If there's a particular point, for instance in epilepsy looking at 30 or 50 percent reduction in seizure frequency, is often part of their endpoint structure.

So the answer is, there are a lot of possibilities, but it has to fit the clinical setting.

DR. FALK: So the right side would be total abstinence of the cumulative responder curves, and that might have been proposed as an a priori primary outcome, and they didn't hit it. Now, if you see the rest of the curves, yes, they're kind of bias because of the differential drop out and everything. So what looks like to be kind of a whopper of a treatment effect somewhere in the middle range -- are you saying that it would be up to them to validate that 50 percent reduction is really clinically meaningful?

FEMALE SPEAKER: So for instance, if you see a large difference in treatment arms in the number of people who have a 30 percent reduction in

Page 318

seizures, who are enrolled with intractable epilepsy, or a 30 percent reduction in pain, or a 30 percent reduction in positive urines, what does that mean?

The epilepsy people are happy. In some clinical settings, the pain people are happy. It doesn't sound like addiction people will be very happy with that outcome based on the conversation we're having unless we decide or you decide that that cut represents a clinical benefit.

So what is the benefit for that population? More people with a 30 percent reduction of the measure, or should it be a 50 percent reduction, or a 90 percent reduction of that measure, what correlates with clinical benefit in that setting?

DR. FALK: Can that be done after?

FEMALE SPEAKER: No.

DR. FALK: That has to be done a priori --

FEMALE SPEAKER: It can be done any time during exploratory work, but --

DR. FALK: Yes, in the phase 2 trial.

FEMALE SPEAKER: -- when you are planning a

Page 319

phase 3 study, I believe the general standard, not just an FDA standard -- but if you want a statistical analysis to be -- well, you folks tell me. How much post hoc analysis do you believe, and when do you call it quits? There's a regulatory reason why we don't like too much post hoc stuff.

DR. FALK: That's what we said. I think we said that in our paper, that maybe this could be -- CPR, the cumulative proportion could be good for phase 2 trials to inform what a good cut point might be for a phase 3.

FEMALE SPEAKER: I'm going to comment that one of the biggest things we learn from these trials is that the cumulative responder analysis sounds like a really great idea. We thought it was going to obviate the need to explore relationships about one particular responder definition. And it's so vulnerable to missing data because there are so many different pieces of data that are collected.

It's so vulnerable to missing data. It's so vulnerable to patients not completing the trial,

Page 320

whether for protocol-specified discontinuation criteria or because for whatever other reason they didn't. And because there are so many unanswered questions about how various places on the curve really predict clinical benefit, that as great an idea as it sounded, it didn't really turn out the way we hoped.

I think we got lucky with the Vivitrol trial. The complete abstinent responder definition also worked, so we did not have to negotiate whether what worked was really a good thing to work. So maybe I would say that we could -- it's easy to adjudicate missing patients if you have one single responder definition. I know people don't like that idea, but it is much less vulnerable to missing data if you can find them and they can tell you that they are using right now. Then you could call them a treatment failure. It seems like it could be less costly to collect that information. This is sort of sexy when you think about it, but then it just didn't work the way we hoped it would. That's my big lesson.

Page 321

DR. STRAIN: Dave, you've been very patiently waiting.
DR. McCANN: I just wanted to comment on the issue of differential dropout, and if you get that, you just have to ignore results of the study. I think sometimes the FDA can be very intelligent and make good decisions, and to acknowledge that remaining in treatment is a benefit compared to dropping out. I think it's one of the best things that they've acknowledged in treatment.
As we do our cocaine trials, we define success very much the way you described for Vivitrol and opiates. To be a success you have to be there in treatment, giving urine, and being clean, and you have to self-report no use. So if you're there and you're giving a dirty urine, or you're saying that you're using, or if you're no longer there and you've disappeared, you're a failure. That can give you an effect by virtue of differential dropout.
I went back and looked at data for a lot of our meth and cocaine trials. I don't do SAS

Page 322

programming, but I asked our statistician, can you print me the last 3 weeks of urine for everybody who dropped out of urine early. I want to see if anybody was clean. I heard it suggested that, oh, sometimes you get people, they're better, they go get a job, and they drop out because they're doing well. I did not see that in the data at all. I didn't see clean people leaving early. It just wasn't there.
So I think that supports the idea of people who've dropped out being treated as failures and requiring people to be there showing clean urines and reporting no use at the end as success. If I had to bet, for Vivitrol and opiate addiction, which way differential dropout would go, I would have expected the people getting that antagonist to say, "Screw this, I'm out of here," and the placebos to stay in.
The fact that you saw it in the other direction to me was really surprising, and it suggested there was something meaningful there, the decreased craving that was shown in those studies,

Page 323

that seemed to be very important.
DR. STRAIN: Ken?
DR. SILVERMAN: This is Ken Silverman again. I agree about retention. Retention is good. And as a measure itself, it's fine and may be an important measure. But I actually don't agree about the fact that you could assume that someone who has left treatment or has not provided urine samples is doing badly, having positive -- your urinalysis results are your outcome measure if you're assuming that they're positive. I just don't think that's necessarily right. It could be right, but it's not necessarily right.
That's all. That's why I was advocating -- and I agree with what Kathy said about just the need to have trials. And if there were these requirements to have not lack of differential retention and high rates of collection, people would find ways to get it done, and it's actually pretty easy. Just like that guy; just pay them. And if you pay them enough --
MALE SPEAKER: The more we pay them, the

Page 324

more likely we're going to get professional subjects who really don't want to quit, but want money to buy drugs.
DR. SILVERMAN: I don't think that's true.
MALE SPEAKER: I do.
DR. STRAIN: George?
DR. WOODY: In the Russian studies, we have followed up people who dropped out, especially looking at HIV risk. And there's a big difference between HIV risk in those that relapsed and those that didn't. There's a lot of data showing such high relapse rates with opiate dependence when they finish detoxification. So I think that in the case of opiates, you can really be pretty sure that maybe not everybody, but a great majority of the dropouts would have relapsed. Now that may not equally apply to cocaine or amphetamines that tend to be less -- you have more episodic use. But with opiate dependence, it tends to have a somewhat different pattern.
DR. STRAIN: Other comments or questions? Connie?

Page 325

DR. WEISNER: I kind of hesitate to say this, but I think a lot of this does have to do with who are in these studies, the whole issue of advertisements. Those people are really different from people who are having to come to treatment because they're in trouble with their job or their family, and maybe you get some of those kind of people. Like I was saying to David, everybody in those Kaiser samples has to take that drug test to somebody, even if it's their wife.

We have really good follow-up rates in our studies because they're there anyway, and they have to be there, or there, for one reason or another, choosing to be there.

In the past, health plans -- I'll talk a little bit about this tomorrow. Health plans haven't really wanted to participate in this kind of research, but it's really different now with the high profile of pain medication, opiates, and marijuana issues that they're dealing with. And they would like medications. So there are different maybe ways, different patient populations

Page 326

that would really help solve some of this. So that speaks for developing relationships with health systems or whatever.

MALE SPEAKER: I think what you're saying is really interesting, and it highlights there are different objectives. There's a component of getting a drug approved demonstrating some sort of benefit. It may sound trite, but it really is the case that that really is just the beginning. I know that colleagues at, for example, the University of Pennsylvania, are doing a lot of work and other places on opiate treatment and the criminal justice system.

So to your point, really getting a drug approved just means, yes, there is some evidence of benefit, but then there's this whole other question about who do you actually use it in to get that benefit and what kind of settings.

Wrap Up – Eric Strain

DR. STRAIN: Not to prolong the conversation unduly, but I want to go back to -- because I think we're evolving into -- or part of our discussion is

Page 327

evolving into a focus upon urine results as an outcome measure that demonstrates abstinence. We could have probably a really quick meeting if we decided that because, all right, we're done. It's here in results you've got to show abstinence. There we go.

But I think part of our charge as well is to think broader in this particular area of stimulant use. We don't necessarily have to discuss it today. It could be something to ponder and to return to tomorrow. But we can think of other things beyond urine results abstinence as the goal, or at least something that we think should be investigated as a potential -- I believe where I'm at now -- so long as we can show that there's a clinical benefit to that.

I don't know what that is, but I think that -- in some ways I think we may be held prisoner by the opioid research stuff, where so much of the opioid research with buprenorphine was people were coming everyday, and we could get urine so frequently, and it was so easy to measure those

Page 328

results. I think we should think broader in this category and be open to the idea that there could be other outcomes beyond urine results that could be identified a research or considered.

I want to put that out there. I think where I'm at is, how do we show clinical benefit? What's the analog to the drink when the drink was used for HDD? Yes, Steve?

DR. SPARENBORG: This is Steve Sparenborg at NIDA. Some people in the room are aware of this, but one small part of NIDA is sponsoring -- through a contract, we're developing a patient-reported outcome for cocaine use. The purpose of this is to come up with a typical kind of questionnaire that is based on, first, expert opinion. Experts have been questioned by professional interviewers from Northwestern University and gotten the lay of the land for what patients are like. And the next step is for interviewers to talk to cocaine users, addicts. And then based on that, go through the typical psychometric process.

We're having the FDA oversee this process.

Page 329

We've just barely started communicating with them. But it will be done through the drug development tool process. And our hope is that although the data shown today don't show a lot of -- it's not really likely that we're going to find a lot based on the lack of concordance with psychosocial outcomes and reduced drug use. But maybe there's something that a cocaine patient could express to the interviewers that would lead us to helping to identify what those issues are and helping the patient to identify with something in the treatment program that they feel more respected, more hope, more something. They feel something differently that we haven't seen yet, and that that could eventually become a lead to something we can study in a trial.

Otherwise, I think it's just leave no stone unturned. Even though the current evidence isn't super promising, we really have to go through this official process, I believe, to say that, all right, we went down this path and this is what we got. It might not be useful, but it just might.

Page 330

Adjournment

DR. STRAIN: Thank you.

If there are no other comments or questions, I want to thank everybody for a real intellectually stimulating day. I think we've sort of gone from soup -- I guess we haven't gone from soup to nuts because that would imply we don't have more to do tomorrow. But we're on the home stretch, and I really do appreciate it.

Just to remind you, we've heard from the perspective of the FDA on this process of outcome measures and their development, and a great talk this morning. Then we've heard from Kyle about the development of a particular measure, some of the psychometrics he worked with. Then we had some great talks from NIAAA about heavy drinking days, and those have really helped me to crystallize in my mind a process that NIAAA has gone through that could be very informative for what's being done here. And then this afternoon we've heard some innovative approaches that have been used by NIDA as they've started putting not just their toes but

Page 331

I think their feet in the water here on this. And clearly, they're very interested in continuing this from the program announcements and FOAs that they've mentioned.

DR. STRAIN: So I want to thank all the speakers today, and I want to thank all of you for your attention, and I hope to see many of you at dinner tonight. Thanks.

(Whereupon, the meeting was adjourned.)

C

D

E

F

I

J

K

M

O

P

S

T

U

V

W