ACTTION SCEPTER II - Clinical Trials to Evaluate Safety Outcomes in Procedural Sedation

---

November 18, 2016

A Matter of Record
(301) 890-4188

ACTTION

SCEPTER II MEETING

Clinical Trials to Evaluate
Safety Outcomes in Procedural Sedation

Friday, November 18, 2016
7:55 a.m. to 4:24 p.m.

Sofitel Washington DC Lafayette Square
Washington, D.C.

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
(7:55 a.m.)
Welcome and Introductions

DR. DWORKIN: Good morning. I'm Bob Dworkin from the University of Rochester. I noticed the slide didn't spell out the acronym, and so I did a little bit of research on the Web. SCEPTER stands for, for those of you who don't have it in front of you, Sedation Consortium on Endpoints and Procedures for Treatment, Education, and Research.

SCEPTER is one of the initiatives that's sponsored by ACTTION. I'm not going to unpack that acronym, and we're all very pleased that you're able to join us for what looks to be a very interesting and important meeting over the next two days.

Could I have the first housekeeping slide? So I'm not going to go through all of this. You can read it for yourself. The most important thing is we all have cell phones and would really appreciate it if you could put your cell phone on vibrate or silence or something like that so that

Page 6

we don't hear your choice of ringtones. The bathrooms are outside.

I guess another important housekeeping item is that we are taping this meeting. ACTTION and SCEPTER therefore are part of a public-private partnership with the U.S. Food and Drug Administration, and we put transcripts of all of our many meetings on the Web so that everything is publicly available and transparent. So just be aware that anything you say for the next two days will end up on the Web in several weeks.

This is the acronym for ACTTION. I just wanted to on behalf of ACTTION welcome you-all and just say a few words about what ACTTION is before turning the meeting over to Dr. Denham Ward.

ACTTION is a public-private partnership with the U.S. Food and Drug Administration. It takes care of what could be thought of as four therapeutic – it covers four therapeutic areas: non-analgesia, pain medicine, anesthesia and sedation; addiction medicine and treatment of addiction disorders; and disease modification; and

Page 7

peripheral neuropathy. This meeting obviously falls within the anesthesia and sedation component of ACTTION.

This is the mission of ACTTION. ACTTION supports with funding from FDA, industry and various other sources, a range of activities. A lot of those activities are focused on optimizing clinical trials, as the slide says, but there are other diverse activities, including developing new measures and outcome measures for clinical trials, developing diagnostic criteria. I'm not going to go into all of that.

I think just to give you a sense of -- this public-private partnership really was the idea of the FDA's, and this, I think, is an informative quote about why the FDA thought public-private partnerships in these areas would be of benefit.

As Janet Woodcock and her colleagues, including Ray Dionne who's here at the meeting, said a number of years ago, "The science base necessary to evaluate and predict safety and efficacy is different from the science that

Page 8

generates the new idea for a drug, biologic, or device."

Dr. Woodcock and Dr. Dionne and their colleagues go on to say that NIH has a history, of course, of funding research in the latter area, basic science research that increases our understanding of mechanisms and targets, and new drugs and devices. But NIH does not have a history of supporting research on the assessment and prediction of efficacy and safety.

The FDA began the ACTTION public-private partnership six years ago now, and a little bit before that, began another public-private partnership that many of you are familiar with, which is Smart Tots. Those are two public-private partnerships that grew out of this view of a gap in what NIH funds and an opportunity of what FDA could support to fill that gap.

That's just a little bit of the background. As I said, we spent a lot of time on clinical trials in all of those therapeutic areas, and there's a lot more information about ACTTION and

Page 9

all of its activities at our website, which is acttion.org.

So unless there are any questions, I'd like to turn the meeting over to Dr. Ward and welcome you-all again.

Denham.

(Applause.)

DR. WARD: Thanks, Bob.

Thanks, everyone, for working with me to help put this meeting together.

Many of you were at the first SCEPTER meeting, but I think not everyone, and it's probably a good idea to go around and reintroduce ourselves because this isn't a meeting where we're going to sit in the audience and listen to somebody come up and pontificate. This is a meeting which we all have to contribute our ideas to reach the goal of how best to look at adverse events in sedation in clinical trials.

I'm Denham Ward. I'm an emeritus professor of anesthesiology at University of Rochester and am now professor of anesthesiology at Tufts, and I'm

Page 10

at Maine Medical Center. We can start with John and go around.

DR. BERKENBOSCH: John Berkenbosch, pediatric critical care at University of Louisville.

DR. CARLSON: Doug Carlson, pediatric hospital medicine and pediatric emergency medicine at Southern Illinois University.

DR. MINER: Jim Miner. I'm emergency medicine at Hennepin County Medical Center in Minneapolis, Minnesota.

DR. DAHAN: Albert Dahan from Leiden in The Netherlands.

DR. CHAPPELL: My name is Phil Chappell. I'm from Pfizer. I work in CNS and drug development.

DR. SEXTON: Anne Sexton, also from Pfizer working in CNS and pain.

MR. WILLIAMS: I'm Mark Williams, and I'm at University of Rochester.

DR. PANDHARIPANDE: Pratik Pandharipande, Vanderbilt University, anesthesia and critical

Page 11

care.

DR. RIKER: Rich Riker from Maine Medical Center, medical critical care and neuro critical care.

DR. WUNSCH: Hannah Wunsch from Sunnybrook Hospital, University of Toronto, intensive care.

DR. BHATT: Maala Bhatt, Children's Hospital of Eastern Ontario, pediatric emergency medicine.

DR. CONSTANT: Isabelle Constant, I work in Paris in children in anesthesiology and infancy care.

DR. ROBACK: Mark Roback, pediatric emergency medicine, University of Minnesota.

DR. GREEN: Steve Green, emergency medicine, Loma Linda University in California.

DR. MASON: Keira Mason, anesthesiologist at Boston Children's.

DR. ZHAO-WONG: Anna Zhao-Wong. I'm from the Maintenance and Support Services Organization.

DR. PETIT-SCOTT: Rene Petit-Scott. I'm with FDA.

AUDIENCE MEMBER: [Indiscernible], clinical

Page 12

reviewer, FDA.

DR. CRISAFI: Leah Crisafi. I'm an anesthesia team leader in FDA's Division of Anesthesia, Analgesia, and Addiction Products.

DR. SESSLER: Dan Sessler. I'm chair of the Department of Outcomes Research at the Cleveland Clinic and director of the Outcomes Research Consortium.

DR. CONWAY: Aaron Conway. I'm a registered nurse from Brisbane, Australia and [indiscernible], Queensland Media Technology.

DR. GOZAL: David Gozal, I'm an anesthesiologist from Jerusalem, Israel.

DR. ROCA: I'm Rigo Roca. I'm deputy director of the Division of Anesthesia, Analgesia, and Addiction Products at the FDA.

DR. URMAN: Rich Urman, anesthesiologist at Brigham Women's Hospital in Boston.

DR. WEISS: My name is Mark Weiss, and I'm an anesthesiologist at the University of Pennsylvania and vice president of the Society of Non-Operating Room Interventionalists and

Page 13

Anesthesiologists.
DR. GAN: TJ Gan. I'm chair of anesthesiology at Stony Brook and also president of the current American Society for Enhanced Recovery.
DR. DEXTER: Franklin Dexter, researcher in anesthesia at the University of Iowa.
DR. KARAN: Suzanne Karan, anesthesiologist at University of Rochester.
DR. LITMAN: Good morning. I'm Ron Litman. I'm a anesthesiologist with the Children's Hospital of Philadelphia and medical director of the Institute for Safe Medication Practices. I'm the ASA's representative here today.
DR. LERMAN: Jerrold Lerman, anesthesiologist in Buffalo.
DR. CLARK: Randy Clark. I'm a pediatric cardiac anesthesiologist working for the University of Colorado at Children's Hospital Colorado. I'm also the ASA's section chair for professional standards, which includes the committees for performance and outcomes measurement and standards and practice parameters, among others.

Page 14

DR. DIONNE: Ray Dionne. I'm a dentist and pharmacologist currently at East Carolina University in the Department of Pharmacology.
DR. IRWIN: Good morning. Mike Irwin. I'm professor of anesthesiology at the University of Hong Kong.
DR. DWORKIN: Bob Dworkin, University of Rochester.
DR. TURK: Dennis Turk, University of Washington, Department of Anesthesiology and Pain Medicine.
DR. O'CONNOR: Bob O'Connor from the University of Virginia. I'm emergency medicine. Good morning.
DR. TOBIN: Joe Tobin, professor emeritus, pediatric anesthesia and critical care, Wake Forest University.
DR. CRAVERO: I'm Joe Cravero. I'm an anesthesiologist from Boston Children's Hospital. I'm the chair of the Pediatric Innovation Research Consortium and on the board of the Society of Pediatrics Innovation.

Page 15

DR. WARD: And Ricky just walked in.
DR. TWERSKY: Hi. Ricky, Rebecca Twersky, anesthesiologist at Memorial Sloan Kettering in New York City. I have been involved with the pre-ACTTION initiatives, and I'm glad to be part of this group today.
DR. WARD: Thanks, everyone. Like I said, a lot of people already know each other, but this reinforces the breadth of expertise that we have here across specialties and across continents.
We tried to organize this as a follow-on to SCEPTER I, which was a meeting where we looked at efficacy. I thought we would start with a review of SCEPTER I.
As I hope you know, many of you are authors on the two papers that came out of that. The first paper was the literature review, the systematic review of efficacy for sedation. The second paper was really the recommendations that came out of the first conference for how you do clinical trials to measure efficacy.
The output of this conference is planned to

Page 16

be a paper on recommendations for how clinical trials should be organized to measure adverse events and how those adverse events should be quantified.
We have a few changes in the schedule, but nothing drastic, so we'll move things along with that. Since everybody knows each other, I'm not going to have any major introductions. We will move from speaker to speaker without any major discussion of who you are.
It's your meeting. So this isn't a meeting to sit and listen to speakers. The speakers, have been working with them, have an introductory discussion, but most of their time should be spent with a discussion from you, which is why there's microphones on your desk. You don't have to get up to go to a microphone. It's all there. We want to get as much input to these ideas as we can during this meeting.
Mark, who was the first author on both our papers, we got him out of call, I think. He was doing vascular cases all day on Monday, and he's

Page 17

going to review what we came up with for the SCEPTER I meeting.

Presentation – Mark Williams

MR. WILLIAMS: Very good. Thank you very much. Thank you, Dan, and thank you, Bob.

It's a pleasure to speak to you today. As we discussed, this is a recap of the meeting that many of you were at a couple of years ago now, so we'll keep this brief so we can press on with the important matters of discussing safety.

This was the overview of the interacting components of sedation and sedation research as we had it in our minds for the last meeting. As you can see, the sedation efficacy and consistency of the center and spreading out. The other important components, we included clinician and patient satisfaction within the efficacy and effectiveness meeting at last meeting. The current meeting is obviously on safety, and I imagine there will be many meetings to follow.

Last meeting was held in D.C. not too far from here in 2014 with 36 attendees across

Page 18

similarly today a range of specialties and a range of adult and pediatric sedation experts, colleagues from the FDA and industry as well.

The overriding impetus was that sedation efficacy is very nebulous concept and consensus on specific outcomes were certainly needed to facilitate clinical trial design and ultimately regulatory evaluations of sedation products.

The meeting consisted of, similar to the setup of this meeting, several presentations, which stimulated discussion following those presentations. A systematic review was not published at that time. The results were available at that meeting, but the article had not yet been published. Some discussion revolved around the results of that review.

The priorities of sedation were felt to be patient and clinician centered with overlapping components of those priorities for this patient and clinician. And reviewing the literature, there were many goals of sedation efficacy of which we touched on: sedation and sedation levels;

Page 19

particularly in pediatrics, behavioral components; satisfaction; sedation timing and procedural-related outcomes, and others such as pain and recall.

We discussed sedation measures as positive evidence for a lot of the sedation measures, which unfortunately not a wealth of psychometric data to support some of the measures. Similarly, in the pediatric sedation scales, we discussed the sedation measures with the most evidence of validity and reliability.

The upshot of the two-day meeting was a paper, which many of you are authors on, which was really built around the domains of the -- we borrowed from the Institute of Medicine -- four of the six domains of the IOM's crossing the quality chasm were used, being safe, effective, patient and family centered, and efficient. For the first paper, we focused on effective and patient and family centered.

Many tools were discussed that could be used to show sedation effectiveness, and we felt that

Page 20

procedural sedation, the procedure's satisfaction really was a universal typical across sedation trials as a way of measuring sedation effectiveness. However, for a drug to be classified as a sedative, we need some form of defining it as having sedation properties. So with that, a sedation scale was vital as well to be included.

Moving on to patient and family centered, the patient satisfaction was considered to be an important aspect of assessing sedation, so that was included in our recommendations. It culminated in the recommended core outcome measures, which we have in front of us.

For the sedation level in adults, the Observer's Assessment of Sedation was recommended for pediatrics, the UMSS. We also included the use of additional rescue medications in there as well.

For proceduralist satisfaction, the clinician's satisfaction of sedation instrument and also observed pain scores as well.

For pediatrics, we had the Children's

Page 21

Hospital of Eastern Ontario Pain Scale and FLACC. For patient and family centered, including patient satisfaction, included two scales, the ISAS and the PSSI, which measure two separate components of satisfaction. They're used independently of each other. And for recall, modified Brice and the Numerical Rating Scale for Pain was considered important.

I think for this meeting, we're hoping to have again presentations and discussion, and hopefully come out with some thoughtful recommendations, which can then lead to publication and further education of the sedation community.

Okay, Denham.

Q&A

DR. WARD: We have some time for discussion on the majors that we have from the first meeting. Obviously, when we were looking at sedatives, effectiveness and safety, obviously, they're closely coupled, and most clinical trials would be looking at both simultaneously.

MR. WILLIAMS: Yes.

Page 22

DR. WARD: Open for discussion and comments from the first meeting. Too early in the morning for anybody to --

(No response.)

DR. WARD: If not, we will keep a little bit ahead of schedule. As opposed to a continuing education meeting, where you want to make sure the talks start on time because people are coming from room to room, we're all in the same place at the same time. If we get ahead of ourselves, that will give us more time for discussion in other areas.

We'll move on to the second set of talks, and Anna is going to be our first speaker to talk on MedDRA and the dictionary for reporting adverse events.

Presentation – Anna Zhao-Wong

DR. ZHAO-WONG: Good morning. Thank you for this opportunity to introduce, MedDRA, the adverse event reporting terminology at this conference. Again, my name is Anna, and I work for MedDRA Maintenance and Support Services Organization.

These are the topics I'm going to go through

Page 23

today. We're going to do the introduction or overview of MedDRA, and talk about what is MedDRA and where is MedDRA used, and who uses MedDRA, and talk about MedDRA's features and how that facilitates adverse event reporting. Then at the end, I'm going to talk about the mappings of MedDRA or integration of MedDRA with other terminologies.

The acronym of MedDRA stands for the Medical Dictionary for Regulatory Activities, and I'd like to do a quick polling. How many of you have heard of MedDRA?

(Show of hands.)

DR. ZHAO-WONG: Well, pretty good. How many of you have used MedDRA?

(Show of hands.)

DR. ZHAO-WONG: I expect some because we have industry colleagues and FDA colleagues. Excellent.

MedDRA was initially created by the International Council for Harmonization -- we call it ICH -- in the early 1990s. ICH, just a quick introduction, is actually right now a legal entity.

Page 24

It's an organization put together by industry and regulators with the goal of setting standard terminologies and best practices so that we can increase efficiency and avoid redundant work.

Because before MedDRA was established, in the world or in terms of adverse events, there were many terminologies used for adverse event reporting. So, for example, in the United States, we used to use COSTART. In Japan, they used to use JART. In Europe, they used to use WHO-ART.

So in other words, for a new drug to be approved on the market by different regulatory authorities, that drug's data or the company who's submitting that drug adverse event data has to be coded in a variety of different terminologies. Although the clinical trial data is the same, but when they submit to different regulatory authorities, they have to code the same adverse event data in different terminology. So that will reduce the speed of the drug approval process and create a lot of redundant work.

ICH was established, and the goal is let's

Page 25

set one standard terminology for everyone to use. That way, we can all communicate. Because in essence, MedDRA is the standard language that we speak in the world of drug safety.

When we're using different terminologies, we're like speaking different languages. Just like if we have a conference, especially a WHO conference, everyone speaks different languages, so they have to have translators so that we can understand each other. But with MedDRA, we all speak the same language so that we can understand. It doesn't matter where you are and to which regulatory authority you submit your data to.

MedDRA is also used in the drug safety monitoring, drug safety communication, drug safety oversight.

With that, we call MedDRA a clinical-validated terminology. It's used by both the regulatory authority and the biopharmaceutical industry, and it's used in data entry, what we call data entry. It's the coding of adverse events. And data retrieval analysis, of course, after the

Page 26

adverse events are standardized, then we can use computer or all the tools they have to analyze the data and retrieve the data.

Evaluation, to analyze is this drug safe for patients to use as part of the drug approval process and for presentation. For example, when the companies submit their new drug application in their adverse event section, they will use MedDRA to present how the drug safety profile is for that particular product.

Now, who or where is MedDRA used? MedDRA is used in the entire product life cycle, including the clinical trials and postmarket, when humans are involved, which means the preclinical. The animal testing stage is excluded. So from clinical phase 1 all the way to the end of that product life cycle, MedDRA is used to monitor and report adverse events.

Naturally, all the regulatory authorities would use MedDRA, especially in their databases, safety databases, and these are some terms that were used in the drug safety world. Like the ICSR,

Page 27

Individual Case Safety Report, would use MedDRA when they do the reporting, and PSUR, period update on the adverse events.

Clinical study reports in the investigator brochures because the investigator brochures will have an adverse event section.

Core company safety information, each company for each product that they will have a master sheet about that product, everything about that product. That's what we call the core company safety information. There's an adverse event section. Of course, MedDRA is used there.

Marketing application for the new drug application. Publications in prescribing information will involve adverse event, and also advertising. There are a lot of patient direct advertising going on, and then on the TV, you will hear the product names and drugs. At the end of the advertising, you will hear they say very fast all the adverse events that may be associated with that product.

Then who uses MedDRA and how MedDRA is used,

Page 28

based on the ICH region? ICH is the organization that initially created MedDRA. And actually when they created MedDRA -- let me back up a little bit. ICH is funded by what we call six parties in three different regions. The three regions are the United States, the European Union, and Japan. Of course, in each region, there are two parties. There's the regulatory authority, and there's an industry association.

The three regions and the six parties funded ICH, and after ICH created MedDRA, then the three regions adopted MedDRA. So the first region is United States. U.S. FDA, although does not mandate the use of MedDRA, U.S. FDA uses MedDRA in its internal databases.

Three FDA safety databases use MedDRA as their adverse event terminology. There's the FAERS for drug and biologics as a CDERS database, and there's VAERS for vaccines as CBERS database. And there's CAERS for foods, supplements, dietary supplements, and cosmetics. So that's for the CAERS database. Essentially, the MedDRA is the

Page 29

de facto standard terminology in the U.S.

Now, in Japan and the European Union, the other two regions within the ICH, MedDRA is mandated for use in the electronic reporting, and of course, we have the biopharmaceutical industry within the ICH regions.

Other than the biopharmaceutical industry and the regulators, we also have MedDRA users in other areas, in other countries beyond the ICH now that more and more countries are adopting MedDRA. For example, in North America, we're looking into Mexico, and Canada already uses MedDRA. In South America, Brazil is looking into use of MedDRA, and in Asia, many Asian countries are doing that as well, for example, South Korea, China, and Singapore, so on and so forth.

Another important use in the MedDRA world is the WHO drug monitoring center, the Uppsala Monitoring Centre. UMC uses MedDRA in its VigiBase so that VigiBase is using the same standards as the regulatory database and industry database elsewhere.

Page 30

We have a large number of academics. We've got universities, research institutes. I would say 20 to 25 percent of all of our users are in that category. We have toxicologists and others.

When we talk about worldwide, we have over 4,000 organizations in our MedDRA community. MedDRA is an organization-based subscription. For example, FDA is counted as one organization, although within FDA, there are thousands, probably tens of thousands of MedDRA users.

Pfizer has headquarters everywhere in this world. They probably five or six headquarters, but Pfizer is counted as one organization in the MedDRA world. One Pfizer subscription is used for all Pfizer staff worldwide.

Next, I'm going to introduce a little bit about the features and structures of MedDRA to see how that works for the adverse events. Now, what MedDRA covers is described on this slide by this big blue circle. Everything within the circle is the information that is covered by MedDRA. Things listed outside of the circle are the ones that

Page 31

MedDRA does not cover.

Let's take a look at the inside circle. Now, because MedDRA is medical terminology, of course, we can expect that MedDRA covers the disease, disorders, the signs, and symptoms. MedDRA also covers the labs, lab tests and test results, and also medical and surgical procedures. And in addition to that, we also cover the patient medical, social, family histories.

In addition to the disease/disorder types of information, MedDRA also includes medication errors, product quality issues, device-related issues, and then pharmacogenetic terms and toxicology-related terms.

Also within MedDRA, there is a unique feature called standardized queries. This is a feature that MedDRA has to facilitate data retrieval and data analysis for drug safety and pharmacovigilance purpose. That's what we cover inside of MedDRA.

Things we do not cover are listed also. I'll start with the top left corner. MedDRA is not

Page 32

a drug dictionary. So when someone is reporting an adverse event related to a drug, they need to identify who is the patient, what type of drug the patient took, and what happened to the patient.

So when identifying what type of drug the patient took, they need to use a drug dictionary to identify the drug, and then use MedDRA to describe what happened to the patient. Did the patient have a headache? Did the patient have a vomiting event or some other events?

MedDRA does not have patient demographic terms. This type of information is captured, but captured in a column that does not use MedDRA to code. MedDRA does not have clinical trial design terms, so in MedDRA, you wouldn't find terms like "double-blindness" or "placebo."

Moving to the right, because MedDRA is also used to not only to report adverse reaction related to drugs but also report adverse events related to drug and device combination products, when trying to identify the device, you need to keep in mind MedDRA is not a device nomenclature. So to

Page 33

identify that particular device, the reporter needs to use a device nomenclature to identify whether it's a pacemaker, glucose pump, or some other device, and then use MedDRA to describe what happened to the patient.

MedDRA does not have a severity descriptor. This surprises a lot of our users at the beginning when MedDRA first came out. There was a why does MedDRA include a severity descriptors? MedDRA has all the adverse event terms, but because the severity of a particular adverse event varies from one clinical trial to another clinical, MedDRA has a standard terminology to use for all clinical trials.

For example, when we talk about a cancer drug trial versus an antibiotic drug trial, the severity will be very different between these two trials. For example, if we both talk about vomiting, the vomiting grade 2 in cancer drug versus a vomiting grade 2 in antibiotic drug trials are very different.

So that's why MedDRA does not have a

Page 34

standardized severity descriptor. That is left for trials to decide, for that particular trial, what is a mild, moderate, and severe for a particular adverse event.

MedDRA does not have numeric value. For those of you who just heard what I said, you said hold on, wait a minute. You just mentioned MedDRA has tests and test results. How come you don't have a numeric value?

The test results in MedDRA are qualitative results; they're not quantitative ones. For example, blood glucose, we have blood glucose normal, abnormal, increase, or decrease, and we do not have blood glucose 40 milligrams per DL or 200 milligrams per DL, so that's the difference. And MedDRA also does not have frequency qualifiers.

What does MedDRA look like? Now, we know the scope of MedDRA, what's in, what's out, so what does it look like? It essentially is a terminology with five different hierarchic level, five tiers. With these five tiers, we can start with a pretty general level called system organ class.

Page 35

You can have system organ class like according to the anatomical body system. You can have cardiac disorders, renal disorders, hepatobiliary disorders, gastrointestinal disorders. You can also have a system organ class based on the physiological system. For example, we have endocrine disorders, metabolism disorders.

We can also have a system organ class based on etiologies. For example, we have an infection system organ class. We have neoplasm system organ class. Then we have an additional system organ class that's not disease and disorder oriented. Like I mentioned in the scope, we have a system organ class for social circumstances for our patients' social and family histories. And we also have a system organ class for investigation for lab tests and test results. These are not disease disorders system organ class.

We also have a system organ class for surgical and medical procedures. So there are a variety of different types of system organ classes, and the total number of system organ classes is 27.

Page 36

Now, with that general topic in mind, on top, when you go down the hierarchy, every level you go down, then that general topic gets divided according to either pathologically, or anatomically, or physiologically, or clinically, whatever makes sense. It gets divided into smaller and smaller groupings. So as the level goes down, the granularity increases.

So by the time you get down to the preferred term level, that becomes a single medical concept. So you could have a system organ class as cardiac disorder, and when you come down to preferred term, we're talking about concepts like bradycardia, arrhythmia, those individual medical concepts. That is what's at the preferred term level.

Under the preferred term, you said, well, we have medical concept, that's done, right? No, we have one more level underneath that. That's called the lowest level term. The purpose of the lowest level term is to provide different expressions of that preferred medical term.

A lot of times one concept can be said in

Page 37

many different ways. That's why our language is so rich. For example, you can have a preferred medical concept called diarrhea. A lot of times in the hospital or doctor's office setting, patients don't usually say "diarrhea," right? They'll say, "I have loose stool, watery stool," all of the other different expressions of the same concept.

That why we have the lowest level term, to allow those different varieties to be incorporated into MedDRA. LLTs can be a synonym to the preferred term or lexical variant to the preferred term. For example, back pain can be also said as pain back. We can have back pain as a preferred term and pain back as a LLT. Then the other types of LLT could be a quasi-synonym or sub-element of that preferred term.

With the different variety of expressions at the LLT level, that facilitates adverse event coding. When patients are reporting different types of expressions, the coder can easily find a corresponding LLT within MedDRA. That's the purpose of LLT, to allow coding adverse events to

Page 38

be linked to MedDRA, and that LLT will lead you to a preferred term, a medical expression.

All of these relationships, the five levels of relationships, are predefined in MedDRA to facilitate a coding presentation and analysis. So when an adverse event is reported to a pharmaceutical company or reported to a regulatory authority, if they have done the coding, what we call the medical MedDRA coding, that means a linked adverse event to a particular LLT.

These are the total 27 system organ classes. As you can see, as I mentioned earlier, it not only has disease and disorder system organ classes, it also has other support system organ classes. So based on the ICH guide, MedDRA is not only used for adverse event reporting, but MedDRA can also be used to encode patient medical histories, surgical medical procedures, as well as lab tests and test results.

That's an example of what a MedDRA hierarchy looks like. This example uses as a cardiac disorder, and it goes down. At the HLGT level, it

Page 39

breaks this cardiac disorder into a smaller grouping, and as you go down the hierarchy, it breaks even smaller, finer group. When it comes down to the PT, it becomes a single medical concept. Underneath that was a different expression.

So now we know the LT is used for coding, and then PT represents the medical concept. What are these three levels for? Those are the three grouping levels to help the subsequent data retrieval and data analysis. So look for safety signals because if you look the opposite way from bottom up, you can tell that similar concepts are grouped together at the HLGT level and then at the HLGT level.

That way with the three levels on top, one can then -- how should I say -- when you look down the hierarchy, we're looking to the microscope, right, to try and find the exact match of adverse event. When we look up the hierarchal level, then we're trying to gather the similar adverse events together. That's when you do analysis. You want

Page 40

to see is there a signal, is there some safety concern that related to this particular drug. Then that's the time that we want to group similar events together, and that's when we want to go up to the hierarchy and to see if there's any particular safety concern.

Because at the PT level, there could be many types of arrhythmia, right? You could have supraventricular arrhythmia, and you could have ventricular-related arrhythmia, conduction disorders. So at the PT level, you may not see a strong signal because different types of arrhythmia are coded to different PTs. But when you move up the hierarchal level, then all the different types of arrhythmia are grouped together, and that's when you start to see a strong signal if that drug really caused arrhythmia type of events.

MedDRA is also translated into many different languages to facilitate the use of MedDRA in non-English-speaking countries. Right now, MedDRA has 11 different languages. English is the master language, and then the English MedDRA is

Page 41

then translated into the other 10 different languages.

All of these different languages are connected through a 8-digit MedDRA code. Each code represents one MedDRA concept, and that concept then in turn is translated into all different other languages. So this workgroup's work will be passed down or adopted by other countries in the world. Now, we start in the United States. Possibly in the future, may be adopted by other countries. MedDRA can then help to link the adverse events to the different languages, other countries that use.

This is the last section I'm going to talk about, the integration of MedDRA with other terminologies. The first example I'm going to use is the CTCAE. CTCAE is an adverse event terminology created and maintained by the National Cancer Institute, and it's used for the cancer trials.

As I mentioned earlier, MedDRA does not include severity descriptors. However, CTCAE, since it's a specialized adverse event terminology

Page 42

and used in one kind of clinical trial -- so the National Cancer Institute does have severity descriptors in the CTCAE.

In their early versions of CTCAE, we created a mapping between CTCAE adverse event terms and MedDRA terms so that CTCAE and MedDRA can be bridged together to facilitate NCI's research, and then facilitate FDA's reporting and drug approval process.

When NCI moved up to CTCAE version 4, what we did is to actually synchronize CTCAE adverse event terms with the exact MedDRA terms. Because all the adverse event terms in CTCAE were in MedDRA. They're just worded slightly different in order to make this bridge easier.

So what NCI decided to do is just adopt MedDRA terms as their adverse event terms. And then NCI, based on the base adverse event terms of those MedDRA terms, defined their grading from grade 1 to grade 5, grade 1 as the most mild adverse event and to grade 5, which is death.

In CTCAE, this base column, the base adverse

Page 43

event terms are MedDRA terms with NCI-defined five different gradings. So as of version 4 of CTCAE, CTCAE is completely compliant with MedDRA.

I should add, CTCAE's terminology and MedDRA's terminology, we both are maintained, and we both evolve further down the road. So we work closely. If CTCAE wants to add a new term to their terminology, they will first look into MedDRA. If their new term that they want to add exists in MedDRA -- if it does, then it's easy to add. If it doesn't, then CTCAE's maintenance organization, the NCI, will contact us, and we can then add that term to MedDRA so that they can add it, and then it's in MedDRA. So the maintenance is important for both terminologies.

This is the TROOPS tool that contains the adverse event terms as well for the sedation purpose, and we have received an initial draft of the TROOPS terms. My colleague Judy Harrison did an initial mapping.

A majority of the TROOPS terms mapped nicely with MedDRA. There's only a handful of terms. For

Page 44

example, here I give the example, the sedation complication, it does not have an exact match in MedDRA. We have anesthesia complication, but not the sedation complication, not at that level of detail. So what we can discuss is to add this term into MedDRA. That way the mapping will be nicely bridged.

By doing the mapping with other terminology also enriches MedDRA because MedDRA is intended to meet the needs of our users. When we did the CTCAE mapping, we added some additional terms to meet the needs of the National Cancer Institute. The last two years, we also did a mapping of MedDRA to pediatric adverse event terminology that was created by the NICHD. In that process, we also added additional pediatric terms to MedDRA, so that through these process of projects, MedDRA is enriched in a particular area of the medicine.

We hope through this process and the collaboration with your terminology, we can make MedDRA better for the sedation society. With that, I'll take any questions.

Page 45

(Applause.)
DR. WARD: Questions?
DR. RIKER: Thank you for that great presentation. One of the things I didn't see in any of your information is the concept of causality. In all of our studies, when we're using MedDRA, there's also a column there, "probably associated, possibly associated." And I think the ability to separate -- just as an example, a varicocele bleeding patient is getting an EGD who gets hypotensive related to blood loss during the procedure. But that's not related to the procedure; that's related to the underlying disorder.
So is there a place in MedDRA for causality to be assessed?
DR. ZHAO-WONG: The causality is just like adverse events. They are disease/disorder terms. MedDRA does not particularly separate these adverse event terms or those causality terms. Since they're all medical terms, what is commonly done is it's in the different fields of the form.

Page 46

I'll give you an example. MedDRA is used to code medical histories. Many medical histories, just like a patient used to have cardiac arrhythmia, that's a history term. And in another, patient reported arrhythmia because he took a drug that caused arrhythmia.
So they're both arrhythmia terms, but if that arrhythmia term is put in the adverse event field in the report, then it's an adverse event. If that arrhythmia term is put in the medical history field, then that's medical history.
I think in causality in your case, the case report form, based on the design, if that disease is in the causality field, then that's a causality. If that disease is in the adverse event field, then it's an adverse event.
Does that make sense? It's linked to the different fields in the report.
DR. WARD: Why don't we wait for the questions? We're going to have a panel discussion with the whole first group, so let's hold the rest of the questions for the panel discussion.

Page 47

DR. ZHAO-WONG: Okay.
DR. WARD: Thank you.
So continuing on this idea of how we classify adverse events, Maala is going to talk about the Quebec guidelines for reporting pediatric sedation.
Presentation – Maala Bhatt
DR. BHATT: Thank you. It's really my pleasure to be here today to talk to you about the Quebec guidelines, which we developed several years ago now. And I don't think I adequately anticipated the diversity of the audience today, so I'd be very happy to take any questions, and I realize that we'll do that in the panel.
Just to give you a little bit of background, we came about this process in anticipation of work that we were going to be leading in Canada through multicenter research looking at the safety of procedural sedation through a long-term surveillance study for adverse events.
Before we embarked on that, we really felt like we needed a standardized list of definitions

Page 48

because in our review of the literature, everybody called the same event different things or they reported different outcomes for the same questions. So as a part of that process, we started out by looking if there were any existing databases that we could use, any taxonomies that we could map to, and really, we didn't come up with anything.
We looked at trying to map our terms to SNOMED CT. We talked to Joe Cravero's group initially to see what they had used, and we really didn't find anything that we were satisfied with.
What the end result was is that PERC, which is Pediatric Emergency Research Canada, who is leading this work, partnered up with PECARN, which is the collaborative emergency research network in the U.S., to develop a consensus panel. I invited Mark Roback to join me as the co-chair on that panel as we had been recently introduced by a mutual colleague. And we assembled a panel of six emergency physicians and two anesthesiologists with equal representation from the U.S. and Canada.
What we came out with was standardized

Page 49

terminology and reporting for adverse events in emergency department procedural sedation. It was a consensus-based process, and I'll just describe a little bit about the process to you and spend more time talking about what we ended up in the formats for our definitions.

The process was, we started off by generating just a complete reference list from the literature from the MEDLINE search from 1950 to the first week of July in 2007 when we started our process. From this list, we drafted a list of sedation terms, adverse events, and definitions found in the reference list articles, and we compiled this and circulated it to the panel members.

Eventually, we reached consensus on the events to be routinely reported, and we did this by way of electronic communication, teleconferencing, and then finally, one face-to-face meeting in Mont Tremblant, Quebec, which is why the guidelines were dubbed the Quebec guidelines.

I'll just describe a little bit to you about

Page 50

how we ended up with intervention-based definitions. In our search, we found that different studies reporting on very similar things reported very different definitions. For example, in one of Mark's studies in 2004 in Denver, so at an altitude, he deemed that oxygen desaturation was a saturation less than 90 percent, no duration specified. Sanborn in 2006 said it was a desaturation greater than or equal to 5 percent from baseline for greater than or equal to 1 minute.

I want you to pause to think about how difficult that is to do in a clinical setting and to see how many of us would actually calculate the 5 percent desat and also wait the 1 minute before intervening. I come from a different lens in emergency medicine. It might be a more realistic thing in anesthesiology, but certainly in emergency medicine, I haven't seen that happen.

Then Dr. Berkenbosch reported in 2004 that desaturation was an O2 sat less than 90 percent for 30 seconds. So what you can see is that all of

Page 51

these definitions have a threshold, a number, and then plus or minus a duration.

When we went rounds and rounds of discussion, we thought that although they are ostensibly very objective, because you have a hard number and another hard number for a level and a duration, there could be two scenarios where you miss these things. If you have a precipitous fall in an oxygen saturation and you intervene immediately, you'll never actually fulfill some of these definitions because you won't wait that 30 seconds or 60 seconds for it.

Then just as I said before, I think duration is a really difficult thing to abide by or measure in a clinical setting where you're really leaping in to help your patient.

We didn't really feel like these definitions were going to be able to give us standardized and reproducible events, which is what led us to this concept of intervention-based definitions. Certainly, I think that they were controversial then, and they still probably are a little

Page 52

controversial. But what they do require is that for both the clinical event to have occurred and for an intervention to be performed with the intent of treating or managing that event. Every event that does occur requires additional documentation.

That helps the researcher. These were developed with the purpose of reporting in research. That helps the researcher sort through accuracy and severity based on the criteria used for recognition and which interventions were performed.

I'm going to go through a couple of examples with you, and that might put this into a little bit of perspective. I'm using oxygen desaturation as the example throughout the next few slides, but certainly, it applies to any of the adverse events.

We defined oxygen desaturation as oxygen desaturation, and one or more of the following interventions are performed with the intention of improving the saturation. The interventions, as you can see, range from very minor interventions such as verbal cues and tactile repositioning to

Page 53

more important interventions such as the application of positive-pressure, ventilation with or without assisted ventilation and intubation.

Then if you do experience a desaturation, we would require additional documentation, and that additional documentation includes for oxygen desaturation, what the baseline saturation was on room air prior to sedation; if the patient was pre-oxygenated; and if they were pre-oxygenated, what method did they receive their oxygen by and what the flow rate was; and then which interventions were performed in response to the oxygen desaturation so that this would allow the researcher or the person sorting through the data to understand for themselves if this would qualify as an important event for them or not. Then finally, what was the lowest reliable oxygen saturation measure during sedation.

I'll use another example here, which is apnea just to give you an idea of another definition. It's the cessation or pause of ventilatory effort, and one of more of the

Page 54

following interventions were performed with the intention of stimulating or assisting with ventilation.

So again, it starts with very mild interventions such as verbal cues and tactile stimulation, but then advances to tracheal intubation and the administration of reversal agents.

The additional documentation here asks the user to indicate the criteria used for recognition. It could be visual confirmation, loss of a waveform. And I think that this really helps the researcher understand if it would qualify as apnea according to them. Then again, which interventions are performed. And we ask them to document all of them that do apply so that we can understand what the most advanced intervention was.

The second thing that we found was that reporting of the adverse events was not standardized. So as we mentioned, the studies that were answering the same question would not report on the same outcomes, and studies used different

Page 55

terminologies to describe the adverse events. Some studies called them type 1 and type 2 adverse events. A lot of studies lumped adverse events altogether even if they had different pathophysiologic origins.

It's a clinically appealing category of airway and respiratory complications, but if you think about all of the things that go into that, such as laryngospasm, partial airway obstruction, central apnea, they all have different pathophysiologies. And lumping them all into one to look at, especially if you're going to look at predictors of these events, I think that you'd be missing some of the granular data.

What we did is we created nine main categories, but we separated the events within each of these categories so that individual events could be reported separately. And if they were lumped altogether, you would have an understanding of what was contained in each of these categories.

For example, some of them only have their one event such as oxygenation, vomiting,

Page 56

aspiration, but others like ventilation contain central apnea, obstructive apnea, and obstructive apnea contains two subcategories of complete airway obstruction and then partial airway obstruction, and then finally, laryngospasm.

You can appreciate that if we just report on ventilatory disorders or a ventilatory adverse event, you really have no idea what's going on with that patient. So I think it really was important for us to separate out those things, especially for emergency department procedural sedation where some of these are more common than others and less common than others.

Then as I said before, each of those adverse events requires supplemental documentation, so documentation that would help the researcher decide on the severity of the event, and as well, the accuracy and what was done to manage the event in some cases.

For example, in vomiting, it's the only definition actually that doesn't require an intervention. So if you vomit, you vomit. It's

Page 57

the expulsion of your gastric contents. But in the additional documentation, we do ask whether an antiemetic was administered to give us an idea of how the event was managed.

Just brief, this is quite short compared to the last one, but I would accept any questions. Just to give you a little bit of reflection, we just completed five years of data collection at six Canadian centers for pediatric procedural sedation, and we gathered about 6300 patients during these five years, looking at the safety of procedural sedation and specifically looking at risk factors for adverse events.

Just reflecting on our definitions, looking at the pros and the cons, I really do feel that the intervention-based definitions give you reproducible, objective events, and that they -- I believe in the intervention. I believe in the intervention over the threshold and duration. It probably would have been a good idea for us to map to MedDRA in retrospect.

I think that another pro is that the

Page 58

researchers can include only events that meet their criteria for severe or important. So we think one of the criticisms was that you captured every event that may not necessarily be important to a clinician or a researcher.

For example, I might have a lower threshold to intervene than a colleague. So if an oxygen saturation decreases to 95 percent, or even 98 percent, I might intervene with a verbal cue, and that would be technically documented as an adverse event, where you might not really think that that's an important event.

By requiring the additional documentation, I would have access to the fact that, okay, only verbal cues were administered, and I would see that the lowest oxygen saturation was 98 percent. So in sorting through the data for research, you could exclude those patients, but an advantage is I guess it's more sensitive, so you don't lose any cases in this way.

The downside, I think, through these five years is that the documentation is really quite

Page 59

extensive. It's six pages of documentation if you're going to document on every adverse event, and I think it really does need to be incorporated into your clinical documentation in order for it to be successful.

We created a site-specific electronic documentation form for each site that incorporated clinical and study documentation into one form, so that this was incorporated into the sedation documentation at each of the sites, and I think that went a long way towards people being compliant with the documentation.

I do think, though, that there is an ongoing need to educate people, the end users, of using these definitions because they really are intervention-based definitions. So just because you need to have the event and perform an intervention -- and the clinical staff did need regular updates when we saw that some of the data coming through was not as we expected, so they did require ongoing education.

That is another thing after Mark's talk this

Page 60

morning, just about the dissemination. With the publication in Anesthesia and Analgesia, I think that the idea is that people across specialties would use these outcomes. But the challenge is -- we published in an emergency medicine journal, and there have been a number of studies that have used the definitions as outcome measures in emergency medicine, but I don't think that these definitions have spanned specialties. So I think that that is a challenge when it depends on where things are published and how things are disseminated.

That's it. Thanks.

(Applause.)

DR. WARD: There's a little change in schedule, and I'm not quite sure exactly how it's going to work. I think we're going to go to common and adverse events in adult sedation, and then Keira and Steve are going to do the SIVA reporting tool in the next session.

DR. PANDHARIPANDE: A little bit in this session.

Page 61

DR. WARD: A little bit of both. So I think these two are going to overlap a little, and then we'll have our question and answer for all the speakers in the first session after your talk.

Presentation – Pratik Pandharipande

DR. PANDHARIPANDE: Good morning. I thought we'd change this format just a little bit to introduce the problems first, and then perhaps the solutions coming from Keira, Mark, and Steve in the follow-up session. So we'll do a brief introduction over here

I'm not going to try to spell out every sedation-related adverse event because that list goes on. Mark and Denham are going to do a review again tomorrow morning on this one, so that's the first part of this.

Quick disclosure over here, I do have a research grant from Hospira, which makes dexmedetomidine, in conjunction with an NIH RO1 that I have.

The important part over here is that I was specifically told that this was supposed to be a

Page 62

discussion, and I'm not supposed to be just presenting slides, so that's going to be the format. It's a discussion format. I'm going to have a few questions, and hopefully, the audience will participate and respond.

I'm an ICU intensivist, anesthesiologist but don't do much procedural sedation. So you-all are the experts out there. I just have to ask the questions. And then the basis of some of these questions come from Keira Mason and Steve Green's work, where they had published in BJA about the reason why one needs to standardize definitions. I'm going to use that as a framework for this discussion.

Here we go. I told you it's going to start with questions. The first important thing, I think, as a group and as we think about recommendations, et cetera, one probably needs to think about what is the definition of a procedural-related or sedation-associated adverse event.

I'm going to just put out the definition that Keira and Steve had put out in their

Page 63

paper -- that's Keira Mason and Steve Green over here -- in their paper in BJA. They looked at the IOM definition. They looked at the WHO definition and sort of came up with this definition, which is more related based on their opinion and their co-authors as far as something that would work well for procedural-related sedation.

I'm going to stop right here, and let you-all look at this and think about this. We can start commenting on whether you feel that this is an appropriate starting point for a definition, or whether this needs to be modified as we think about what our recommendations are going to be for other folks. We'll start with Rebecca.

DR. TWERSKY: I guess my reaction is to the first word, "unexpected." We know that when we give sedatives and analgesics, that we're going to have some sort of respiratory response whether it's apnea or a delay in respiratory rate. So I wouldn't necessarily consider that unexpected.

I think it is an adverse event if, again, we come up with a definition of duration in the

Page 64

intervention. But I think we might consider responses to the sedation that are adverse but are not necessarily unexpected.

DR. WEISS: At the same time, to follow up on what Rebecca was saying, that "cause or threatened to cause." For example, if you're in a GS, and you give propofol, if someone becomes apneic for seconds, might slip their jaw for a second when they breathe, that threatens to cause an adverse event. But I don't feel that -- and I'd intervene by definition. But then once I intervene, and they breathe again, I don't consider that to be an adverse event.

I'm wondering if there's an issue of sensitivity and specificity. Are we capturing too much? Are we capturing things that may not make a difference, and then might alter the way we treat patients, when in fact we have a hair trigger on what we call an adverse event?

I don't mean to sound cavalier about that, but if I put my finger on someone's jaw, and they breathe 2 seconds later, I don't consider that an

Page 65

adverse event.
DR. PANDHARIPANDE: Keira?
DR. MASON: I don't quite agree with Rebecca about the respiratory depression or whatever, that when we're doing a sedation, that we necessarily expect that we're going to have an adverse event. I think the opposite: When we do a sedation we don't expect, we're going to have an adverse event. We anticipate that we will have some events. Maybe we might have some respiratory changes.
Certainly, there are drugs that don't even create respiratory changes, and you might have a hemodynamic change. But I think it's what you don't anticipate is going to happen that we are really trying to capture.
DR. PANDHARIPANDE: Maala?
DR. BHATT: I was just going to make a comment. I think that that indicates that you are very high-skilled. So if this is to be adopted by everybody, the small events are precursors to bigger events.
So if we don't recognize them, if we don't

Page 66

see them as a -- you don't have to call an adverse event, but if you don't see it as a complication or a precursor to be a bigger event, I think that people that are less well trained, or aren't anesthesiologists, or practicing in a small community where you might not do this as much, may not see this as -- might not view it in the same ways.
Does that make sense?
DR. PANDHARIPANDE: Joe?
DR. CRAVERO: I think this is all really good work. I would just offer a couple of thoughts of what we've talked about in our consortium for a long time, which is it is hard to make definitions that fit every type of provider, because as an anesthesiologist, I may be providing what I'm terming sedation with propofol.
Honestly, if a patient becomes apneic for a period of some seconds during that case, almost like the Geico commercial, for me, positive-pressure ventilation is what I do, so I don't necessarily consider that an adverse event.

Page 67

Even if there was apnea and I intervened with positive pressure, it's part of my work every day.
I would offer -- and again, I'm not saying this is correct, but this is the kind of conversations that come up. That same care delivery in a different setting with a much different provider, or that same event that occurs with an oral sedative having been delivered by a nurse provider, where the patient becomes apneic and there is a requirement for positive -pressure ventilation is a slightly different situation.
I would also offer the kinds of things we've talked about, which is these minor issues like oxygen desaturations, that we are assuming are precursors or harbingers of other bad events, there are not a lot of papers that really help us understand what a 10-second oxygen desaturation less than 90 really means in terms of any kind of outcome. In and of itself, it clearly doesn't represent an adverse outcome.
Whether or not events like that are actually connected to more severe outcomes, there are people

Page 68

in this room that are better outcomes and statistical researchers than I am. But that connection is not necessarily made in most settings of sedation.
So I would just say while I think they're important in one sense or another, we do need to be careful about how generalized some of these -- at least when you get into the weeds, it starts to get kind of tricky.
DR. WEISS: Let me ask you a question then. Are you saying then that leads to the possibility of raising the idea that what might be considered an adverse event in one setting is not an adverse event -- with the exact same set of situations, what might be an adverse event in setting A is just not routine but not unexpected and not an adverse event in another setting.
So it's not just the adverse event we're dealing with, but the location and site with which we're doing it that might also provide the destination, what we're going after.
DR. PANDHARIPANDE: We'll let Doug respond,

Page 69

and then Dan Sessler.
DR. CARLSON: When we look at existing taxonomies of patient safety events, most are based on outcomes -- and I think if we look at this, it may help a little bit, although this is the crux that gets to be difficult -- is that if you look at serious safety events, or temporary or permanent harm, you go back to whether there was a variation from standard care. And there has to be a variation in standard care to actually go into a safety event.
Now, I agree that bad outcomes in sedation are always variations of standard care, but it gets back to apnea. If you have an apneic event and you are trained to do that or are expecting that, that's not a variation from standard care. So I think we have to be a little bit careful about saying that is the adverse event.
On the other hand, I do think that those interventions should be proxies for precursor or potential near misses. It's balancing that of measuring all the things we do to intervene versus

Page 70

what is not standard care and separating it out that is the crux of the issue.
If we look at outcomes and go backwards, there may be a solution, although not an easy one that I see.
DR. PANDHARIPANDE: Dan Sessler?
DR. SESSLER: In the previous version of this meeting, we had a problem in that events were considered to be serious or not on a highly contextual basis. For example, movement in some situation was considered absolutely fine as long as analgesia was okay. In other situations such as pediatric MRI, movement was a disaster, but you had no need to deal with amnesia.
The way we got around that was making our primary outcome based on proceduralist satisfaction. And I wonder if we do something similar here, where complications are defined in terms of the context and who is performing it. Complication would be something that the proceduralist considers to be abnormal.
An anesthesiologist is giving a little

Page 71

positive-pressure ventilation is not considered abnormal. That's absolutely fine. It's a nurse who is unprepared for this in a different context, maybe that is an adverse event.
DR. PANDHARIPANDE: Training plus the context of the --
DR. SESSLER: Exactly.
DR. PANDHARIPANDE: We have one here, and then there, and then we'll get to you. Sorry.
John Guerra? Sorry.
DR. GUERRA: I think sometimes we get hung up a little on event versus adverse event, and two aren't necessarily the same. As an intensivist, I may be providing positive pressure as well during procedural sedation. That's okay. That's part of what I'm trained to do as well.
Might I call that an adverse event? Maybe yes, maybe no. But at the same time, picking up those events, even if they don't lead to a patient outcome that is a problem, is important because it helps us in defining something that we haven't discussed yet, and that is, what's the skill set

Page 72

required at the bedside when a patient is undergoing procedural sedation.
I think there's value in collecting both of those, and we can argue back and forth probably about expected/unexpected, adverse event/event, but yet defining those things helps us become probably safer sedation providers in the long term.
DR. HERTZ: Also, what if you have two agents, and they're both resulting in these events that are readily managed, but one is doing it at twice the frequency? I think that's something that people would want to know when they're selecting an agent, what is the difference, and if you can't capture these things in some way, even if they are expected, how do you make a judgment about the overall utility, all of the different decisions that are made?
DR. PANDHARIPANDE: Denham, and then Keira.
DR. WARD: This is a great discussion, and I think you also want to think about, even though there's a lot of overlap, maybe importing as a QI system, where we're letting a lot of practitioners

Page 73

use a lot of different drugs that we want to keep track of, maybe more like a postmarketing QI situation, versus if we're designing a clinical trial for a new agent in a phase 2/phase 3 type of trial, how do we define adverse events prospectively so we're collecting that data for the approval process in a phase 3 clinical trial.

There's a lot of overlap there, but they're somewhat different, too, in the kinds of adverse events that we're going to be looking at because in the QI situation, we're much less controlled, right? We're going to be in different areas with different practitioners doing different kinds of administration, versus a phase 3 clinical trial, it's going to be much more controlled: who's going to be given the drug, how we're going to be collecting the data, what kind of situations. It's going to be in and, perhaps a lot more control over the kinds of definitions of adverse events that we're going to be able to collect.

Maybe, overlapping in the two concepts of an adverse event and a QI type situation and adverse

Page 74

event in maybe a new drug, or a new technique, or new device. We can discuss both, but I think we want to keep the focus a little bit on the phase 3/2, maybe even phase 1, clinical trial, that I'll talk about tomorrow, of a new drug or device going through the regulatory process.

DR. PANDHARIPANDE: Thank you. Keira?

DR. MASON: I think when we're trying to think about what is an adverse event, that maybe we need to define what is sedation because my definition of sedation is patient who is able to maintain hemodynamic stability, maintain their own airway on their own.

If that's what we're defining sedation as, then any time somebody is doing positive-pressure ventilation of any kind is a deviation from what essentially the definition of sedation is. I think it's irrelevant whether I feel comfortable ambuing a patient because that's my skill set as an anesthesiologist. That is not necessarily the goal of what sedation is, so it's a deviation.

AUDIENCE MEMBER: So I was just looking at

Page 75

the wording of the definition. Is it possible to put the first slide back up, the first question? I didn't realize we were on 3 already.

DR. PANDHARIPANDE: Well, the group discussion went longer than I'd anticipated.

(Laughter.)

AUDIENCE MEMBER: Whenever I look at these things, I try and get rid of terms that you can't really define that are too vague. Although I agree with pretty much everyone's -- what they've said, you can look at these kinds of definitions and say, well, what would "unexpected" actually mean or "undesirable"?

I would get rid of terms like that or even the word "threaten," but I would combine just simple facts like "responses that cause patient injury," I think we can all that most people know what discomfort means. And then Dan's recommendation about the provider contextual is great, and to combine those two things.

DR. PANDHARIPANDE: Rich?

DR. RIKER: I think if we think about the

Page 76

variety of patients, procedures, and adverse events we're talking about, it's incredibly complex to try to pull something that's going to apply across the board. But I would really plead for us to have the ability to understand what was sedation related and what was either disease related or procedure related. A patient gets intubated during bronchoscopy, that might be an incompetent proceduralist causing pneumothorax. That might be over-sedation and apnea and needing intubation for that. That might be an underlying disease process, where the patient was on 80 percent oxygen but not intubated prior to the procedure.

So having some ability to make sense of that and assign that etiology to the adverse event I think is another thing we really need.

DR. PANDHARIPANDE: Sure. Last person. Mark?

DR. ROBACK: I think as we identify these events, or adverse events, or adverse outcomes, we need to consider what we're going to do with that information at the end of the day.

Page 77

When Maala first presented her Quebec guidelines to Joe's group, we had this vigorous discussion, which was exactly like we're doing now, and it became very clear to us -- because in emergency medicine, we do sedation, but we don't necessarily do it every day, whereas the people that are doing it every day in their sedation units, they're going to be measured by their outcomes and their adverse events. So every time Joe does a jaw thrust, they're going to say that's a bad thing? Well, of course, we don't want that.

So really considering what are the most important things to follow and with patient safety being the goal.

DR. PANDHARIPANDE: All right. I'm going to move on to another question. This is just to get the discussion going, which I see we've gotten that goal taken care of.

(Laughter.)

DR. PANDHARIPANDE: This question leads to the next two presentations, which are going to be talking about the tools. We've already started

Page 78

introducing this concept that is it just I expect some apnea, I expect some of this.

So when we think about definitions and recommending what definitions should be considered adverse event, should they be linked to events and thresholds? So you had an apnea period for X amount of seconds, or does it have to include an intervention?

I'm just going to put up a couple of examples over here. All of you know this, but these are from the literature, in apnea for 30 seconds or oxygen saturations less than 90 percent for 30 seconds. These kinds of numbers, they have disadvantages because there are no thresholds based on the fact that nobody has been able to show that this particular thing is associated with an outcome, which is some of the things that we've been discussing now.

That's one way of doing it, is having event threshold base, and Maala has already discussed some of this about having an intervention-based definition. For example, would you consider apnea only to be an adverse event if

Page 79

you required masking or positive-pressure support, or an oxygen desaturation would be considered a true desaturation only if you required oxygen supplementation.

Again, those seem to have some benefits, but there are some problems as well. Because if these are to be reported, do you think someone is not going to be reporting something because they don't want it to be an adverse event. So if I can get by -- I see the sats are now 89, I see they're 88, 87. They will recover. Let me just give them a little bit more time, so those kinds of things and whether that causes a problem.

I'm going to again open it up for questions because I don't want to show the scale yet.

TJ?

DR. GAN: So again, as you alluded to, the problem with that is that we all practice differently. We have different anxiety levels of when to intervene. One may intervene when saturation is 95 percent; others may intervene at a different level. So then you end up with a not

Page 80

very useful data because everyone intervened at a different time when you start intervening with blood pressure going down by how much for how long.

So I think it's important to perhaps capture the raw data, so to speak, when the saturation drops X amount or blood pressure drops an amount. Then whether you intervene or not, that is again, as Dan has alluded to, is contextual. Some people intervene -- an anesthesiologist may intervene at a different level compared to the others.

I think the problem with this, what you put up, is that it's going to be very difficult to sort out what the actual events mean.

DR. PANDHARIPANDE: I know there were problems with what I put up. That was the whole reason I put it up, to start this conversation.

We'll go next there, and then, Maala, you're next.

DR. LERMAN: I agree exactly with TJ. I think the construct in which you're making your observation makes a difference. So I think you need to capture both groups of information. I

Page 81

think however arbitrary your initial thresholds for identification of a "adverse event" may be is one thing, an intervention suggests an increased level of concern, and that raises the bar. You could call it major to minor or otherwise.

For example, who in the audience would not intervene if the patient's saturation were 80 percent? So it's pretty obvious, we're using the 90 percent and below as just a buffer because the next situation may become extremely concerning. And if it gets to 80 percent, if you didn't intervene, with a bradycardia, for example, you almost certainly will be running into a problem shortly.

It is totally arbitrary. It totally depends on the individual and the construct in which this occurs, and I think you need to capture both bits of information. Individually, I don't think you can ever come to a satisfactory conclusion about what an adverse event is.

DR. PANDHARIPANDE: Maala?

DR. BHATT: Obviously, I have an inherent

Page 82

bias here, but I would agree that just having a definition that includes an intervention and classifying that, and lumping it all together, is not useful information. But I think that I would agree that you need both sets of information. And I think that the required documentation that follows in the intervention-based definition will provide you with that.

I would still maintain that if you use the thresholds, I think that you're not going to get an accurate, reproducible event because different people will intervene for different things. And just because they are part of the study, I don't believe that they will stand by and wait for that threshold to become effective.

DR. PANDHARIPANDE: We'll do Rick, Denham, and then John, and then the next speakers can come up with some solutions.

DR. WARD: The next group is a panel, and maybe we should take that right now, is to get the speakers from the first session all up here. Maala and Anna, you will be on the panel, and we'll just

Page 83

segue right into the panel.

DR. PANDHARIPANDE: Do you want them to show the tool at least and then walk through the world -- Mark, do you guys want to introduce the tool and then -- do you mind that?

DR. DENHAM: Yes, briefly, because I want to make sure we leave enough time for the panel discussion.

DR. GREEN: Pratik, should we present the World SIVA, the previous tool, but we'll wait for the new one until after?

DR. PANDHARIPANDE: The next session, yes. I think that might work out and give people time.

DR. WARD: We have a larger panel. So we're doing the previous and the new too, correct?

Presentation – Keira Mason

DR. MASON: Steve and I actually worked on this adverse event sedation reporting tool when I was chair of the International Sedation Task Force for the World SIVA.

What we were doing was trying to address the problem that's already clearly been stated, that

Page 84

the challenge is defining sedation-related adverse events, defining what the meaning of it is, and also what the potential implications of these events were. As we all know, when you read the sedation literature, it's multi-specialty involved, both adults, both children from all parts of the world, both developed and developing areas of the world.

The challenge is looking at the way that the data was collected, the content of the data, the definitions that were used to describe the events, the interpretation of the events, and of course, then what do they mean in the context.

Our goal was to come up with a standardized set of definitions, originally, for the sedation-related adverse events. The initiative that Steve and I are here to talk about was, of course, the World SIVA, which is the adverse event sedation reporting tool, AE sedation reporting tool, and then Mark's going to come up later and talk about the evolution of the World SIVA tool into the TROOPs, which we will talk about.

Page 85

The World SIVA International Sedation Task Force consisted of 25 physicians from 10 specialties from 11 countries both adult- and pediatric-focused clinicians. They had to be not only doing sedation in their daily practice and/or also -- but definitely involved in sedation-related research.

We really had quite a collection of expertise, some of whom who are actually in this room today. We had a group meeting, multiple correspondences in terms of emails, in terms of trying to come up with and agree on these definitions of adverse events.

As you can imagine, it was very challenging because we had everyone from gastroenterologists who do just adults to the anesthesiologist who is overseeing literally technicians providing sedation in areas of Africa where there were no physicians or expert providers at all.

What we came up with was published in the British Journal of Anesthesia a few years ago. It was the "adverse event reporting tool to

Page 86

standardize the reporting and tracking of adverse events during procedural sedation."

Just briefly, I think the strength of this tool was that not only did we come up with agreed upon definitions for these adverse events, but again, beyond defining these adverse events, what were the interventions, and then also what was the potential risks involved and the outcomes of these interventions.

Then at the end, we came up with a descriptor of what was the outcome. Was it a sentinel outcome that had significant adverse events, or was it something that was just very, very minor and transient? Again, at the end of the tool, which had six parts, we had everything from a sentinel event, to a moderate event, to a minor to a minimal event.

The format of this tool was an evolution of the Quebec guidelines that Maala presented because we did go into the actual interventions that were needed to be performed.

I received an unrestricted educational grant

Page 87

from Hospira, and we actually put this on the Web. So this is an open access Web-based tool. It's meant for anybody in any area of the world. What I liked about this is that there were no HIPAA identifiers. But also for those who are in areas of the world where they aren't able to collect their sedation data in a standardized fashion or an organized fashion, they could with their user name and password be able to collect and pull up their data at any time. And especially for people who are -- like I was called from Saudi Arabia because they failed their International Joint Commission visit for sedation, they could potentially be using this to start tracking their adverse events.

There were challenges. Nothing is perfect, so one of the challenges that we saw that evolved into our new project, which was TROOPS and the formation of the new committee, which was called ICAPS, the International Committee for the Advancement of Procedural Sedation, it was based on our identifying that not all of the adverse events really were reflective of the outcomes, and that

Page 88

were certainly challenges, some of them because we hadn't necessarily organized by organ system.

Some people felt that they were doing the sedation tool for tracking and identifying minimal risk outcomes, which might not have necessarily been time valuable for them, and again, that there were some problems, like Maala had already mentioned and others, with identifying the thresholds.

For example, an oxygen desaturation, we couldn't agree. If you're sedating a patient in the cardiac cath lab who's already coming in with an oxygen saturation of 75, what is their desaturation going to be identified as, and for how long would that need to occur for it to be identified as an adverse event?

That was pretty much what we worked on for the AE sedation reporting tool.

Steve, do you have anything you want to add?

Presentation – Steve Green

DR. GREEN: Yes. I just want to add that last point about the thresholds and duration, a lot

Page 89

of feedback that we would get is everyone has a different idea of what the threshold should be or what the duration should be. So to incorporate or to continue with some kind of definition, you're guaranteeing that people are not going to be able to agree on it over time.

Q&A and Panel Discussion

DR. WARD: Can we have all the speakers up from the first session? You guys, too.

We can run over a little bit because I think the session next time is going to be a little bit shorter.

DR. PANDHARIPANDE: The TROOPS can be about 20 minutes.

DR. WARD: For TROOPS, yes. So we can go a little bit longer.

The ideas that I come away with so far is we do have some tools out there for classification of adverse events. From my perspective, they're a little more aimed at the QI situation where we have a lot of practitioners doing different things, and less towards the clinical trials situation where

Page 90

maybe we can specify the threshold and the duration for the intervention in the design of the clinical trial and specify what the signal is that keys the intervention, that reporting the signal is important, not just the intervention but actually was it a saturation? Was it the patient reporting nausea before they actually vomited as part of the signal for giving the ondansetron as an intervention?

I think there are some issues that we've got some tools, but are the right tools and how do we modify them if need be for the clinical trial kinds of situation?

Opening it up for the panel and for continuing the discussion that we've been having.

Mark?

MR. WILLIAMS: Just talking about provider, the thresholds can be very provider specific. One thing I've seen is what do people think about having time outside of a specified threshold as an outcome?

DR. BHATT: How is that different from

Page 91

duration?

MR. WILLIAMS: It can be provider specific, so you can set that. It can go across different specialties, their parameters. But maybe multiple times throughout a sedation procedure, you can dip below 92, 90 for 15 seconds, 30 seconds.

DR. WARD: Instead of one 30-second period, maybe out of the 20-minute sedation, you've had several segments of desaturations, none of which lasted 30 seconds, but in total saw the area-under-the-curve kind of concept.

MR. WILLIAMS: Does it matter more? Does it not matter? Just a thought, a suggestion. Value your opinions.

DR. PANDHARIPANDE: I just feel that it gets more complicated if someone has to measure the area under the curve where someone is out of the threshold. As a reporting tool where you're saying it's something that has to go across specialties, across nations, I think there are challenges associated with that.

MR. WILLIAMS: Certainly, across certain

Page 92

countries -- our institution, we have electronic reporting. All our data is grabbed in from -- especially the saturations, it's captured every minute, so there might be a way of capturing it much more frequently than that.

DR. WARD: Remember, clinical trials may be different than a reporting tool in a QI situation.

I think Albert had a question here and then -- Albert?

DR. DAHAN: In my research and focusing on saturation is not really my aim. Saturation is not the endpoint of -- or maybe it's an endpoint. It's not the cause of the adverse event. The adverse event actually is the patient is not breathing well enough, and how do you cope with that is much more important than looking at saturation. It's much more complex than just breathing. It's a measure of gas exchange.

So we are looking at actually breathing, especially pattern breathing of the patient. It's not very difficult to measure, but it takes some training, takes some time.

Page 93

That's what we're doing currently. We're looking at especially the variability of breathing. If variability goes up, believe me, within a couple of seconds or minutes, the subject patient might stop breathing.

We're really much too much focusing on endpoints rather on cause of the adverse event, in my opinion.

DR. WARD: Any comments from the back or -- Dan and then John.

DR. SESSLER: I guess one of the challenges we face here is that we essentially do not have a link between observed events and outcomes. In that respect, it differs from blood pressure where we now know what the association is between different levels of hypotension and outcome and can evaluate those associations across a variety of different measures.

One paper that evaluated measures of hypotension that have been reported, they found 140 different measures reported in 130 papers. This is not really very helpful, but I guess I see the big

Page 94

problem here as a lack of link between events and outcomes. We're saying that these are events that might foreshadow problems and that if you don't do anything about hypoxemia, eventually, you will get into trouble, but we don't actually know where to intervene.

I guess that brings me back to proceduralist or sedationist and context as being really important because what's an important event in one context may really be completely unimportant in another, and the danger is that we record a bunch of events. It's technically easy to record events. You can record every episode of desaturation, and you can do more sophisticated things like area under the curve or time-weighted average below some threshold. But we still don't know what it means, and what it means is going to depend very much on who's there. That's especially true when you get to interventions because an intervention that's trivial for an anesthesiologist may not be in another context.

DR. WARD: In the context of a clinical

Page 95

trial for a new compound, would that then -- we will talk about some breakthrough issues later on. Would that then change the indications and usage? Like when propofol first came out, who could use the drug based on the data that we got from clinical trials?

DR. SESSLER: Right. Well, we have the FDA people here who can comment, but I would assume that if you're testing a new drug that the results apply in context and the FDA labeling may reflect that. But maybe you could help us, Leah.

DR. CRISAFI: I'll let Rigo go ahead.

DR. ROCA: This is Rigo Roca, and actually Dr. Hertz is back there as well. We agree in the context that when you get the data, you're able to actually try to get a picture of what the safety profile actually is and whether there are certain events, as has been discussed before, that really do not require a lot of intervention. That's actually useful to know.

As Dr. Hertz mentioned, we would be able to have information regarding the potential

Page 96

comparisons with different drugs, et cetera, and that information we would try to put into the package insert to inform you so that you know what was seen in the clinical trial.

DR. WARD: John?

DR. BERKENBOSCH: I have comments and then a question. First, I'm going to just say and I think that there's little value in differentiating events based on provider specialty. I think that's unhelpful. It's divisive and probably not constructive to advancing sedation-related clinical trials.

The question I had for you, Maala, using the Quebec guidelines, and there's a lot of value in looking at the intervention part of it. What do you do with all of the data that isn't collected, that isn't recorded where maybe somebody's hypotensive for a period of time, and the provider thought, nah, I don't need to intervene because the other ones look okay? I think that's still potentially valuable data.

What do you do with that in the setting of

Page 97

reporting using the guidelines, or any intervention-based guidelines, whichever one you choose to use?

DR. BHATT: That's fine. I think that maybe somebody else could chime in because my answer for that is that we actually don't do anything with that data.

We have a number -- so with propofol, if they have a transient drop in blood pressure and the practitioner doesn't feel the need to intervene, we don't actually capture that data. Because the thinking behind it was that if it is a significant event, that there will be the need for an intervention. You can't have hypotension that gets worse and worse and worse without an intervention, right?

So we don't have that data, and we don't have -- I think what Mark was alluding to is the electronic capture of vital signs that get stored. Certainly, we don't have that at our center or any of the centers that we worked at, but that could be useful information with that respect.

Page 98

I'd be interested to hear what other people think about, that thinking, because I think that there is -- I definitely come from one way of thinking, and there is a disconnect with understanding what to do with that data or how people feel about that. I'd be interested in hearing what others think.

DR. RIKER: As we look at the individual adverse events, we could come up with specific interventions that might be a long list and would vary by adverse event. But I wonder if a simpler method might allow us to allow more flexibility. I think about a rescue event like a jaw thrust or a few breaths with a bag-valve mask or something like that versus something that extends beyond the procedure.

You go to the ICU, you get intubated, you're on new antibiotics, something like that as just a measure of what might be a minor or a simpler event versus something that extends beyond the procedure and requires a higher level of care or something like that.

Page 99

DR. WARD: I think what we're hearing here is, again, the reporting for a QI situation versus a clinical trial situation may be somewhat different, and at the different level of clinical trial, do you need different levels of data?

Phase 2 trial, you really want to know all the saturation data and maybe not -- maybe as Albert pointed, saturation is too far down the line. You really want to know more about the actual ventilation. Saturation is actually a fairly difficult parameter to measure.

Like Rick was saying, does the severe outcomes, somebody gets admitted to the ICU because they vomit and aspirate, that's clearly an adverse event. But in a phase 2 trial, what are the kinds of things that you're going to want to be collecting there as opposed to a phase 3/phase 4 clinical trial?

Anybody on the panel?

DR. GREEN: I'll just weigh in. I think a lot of this discussion about clinical QI is very relevant to FDA because first we're deciding what

Page 100

are the things that are clinically important. Then there may be another layer of data collection below that that's needed for a phase 2 clinical trial, but I think the first discussion tells you what are the most important things that the end users are going to care about.

DR. WARD: I think we get to the problem that Dan has alluded to. We can get what the adverse events are, but do we know what the signal is in more of the physiological data that would be predictive of it? We may know that for some of the work that he's done in blood pressure. We may not know that in some of the other possible adverse events.

Ricky?

DR. TWERSKY: I think what would help is we have on the dais panelists who have knowledge about the registries that we've collected, and Joe Cravero, and maybe you're going to be doing that later. But I think what would help me in understanding how we fill out these ambiguities by learning about the robust information that has

Page 101

already been collected, again, it wasn't in the clinical trials; it was in the context of clinical care.

But to help us then narrow down these questions that have been brought up as far as duration, level, hypotension, hypertension, I'd like to hear -- and you don't have the slides up there, but that would help also to inform us what you've seen from thousands of cases that you've looked at.

DR. CRAVERO: If I can just say, Rebecca, I will overwhelm you with slides.

DR. TWERSKY: Can't wait.

DR. CRAVERO: Minutiae detail on what we found. And I think it does inform this conversation a little bit, but the exact issues that are being brought up here, I don't think are changed hugely, that you have a large number of very minor things that are reported and a very small number of very major things reported in the pediatric databases.

Like I said, I'll show you examples of our

Page 102

data, but I think the issues remain difficult in terms of exactly what everyone has been saying here: what represents an outcome versus a complication? We have spent a lot of hours discussing that, and I think this is a good conversation. But we're running into the same things that we've done when we tried to come up with a consortium reporting tool.

DR. TWERSKY: Right, because I don't think we'd want to be bogged down with minor events, and that could be what's happening in your reporter registries, or if you had the same experience.

Dr. Bhatt?

DR. BHATT: We are just about to publish our first paper. Hopefully, I'll submit it while I'm here at this conference. We separated things that we didn't -- we reported on four major outcomes: serious adverse events, adverse events that require significant interventions, oxygen desaturation, and vomiting because they were the most common things.

I think that when it's from an emergency department perspective, it is more clinical, and I

Page 103

appreciate what Denham is saying with that. I do think that the things that we report on, they're not going to be the same things that you want to report on in a phase 2 clinical trial.

I think that it's fair that you're going to want much more granular, different information, and I think that that's worth pursuing in terms of what to report and how to capture that data. But I would still maintain that I don't think that the threshold duration is the answer there.

I think that there is another answer, but I don't know what it is. But I don't think it's threshold of duration.

DR. WARD: Other comments, anybody else want to weigh in?

DR. ZHAO-WONG: When we talk about adverse event definition, we need to keep in mind adverse event versus adverse action. Adverse event is actually an undesirable event regardless of causality.

DR. WARD: We talk about more granularity of the data because in a clinical trial, the more

Page 104

granularity you have, the more expensive it is, too. I think it's nice to collect everything, but that gets more and more expensive to collect everything in a clinical trial.

You'd like to collect granularity of things that are going to affect -- and that gets to what Ricky was asking. What are the outcomes that are actually occurring, and can we collect data earlier in clinical trials that are going to be related to the actual outcomes that we see in these QI databases?

Hannah?

DR. WUNSCH: Just a comment on hoping that looking at the long-term outcomes maybe answers some of those questions. As someone who does a lot of work on mechanical ventilation, we always talk about patients who receive mechanical ventilation, not require mechanical ventilation, and are admitted to ICU not requiring intensive care for the exact same reasons we're talking about, the small adverse event category, you get the exact same problem when you go to the next level, even

Page 105

though it feels like maybe it should be a more concrete answer to some of these questions. I think you just dive into the same problems.

DR. KARAN: I'd like to echo some stuff that Dan was saying and that Albert was saying, is that we're actually just not monitoring ventilation right now in procedural settings. It's very hard to assess what's happening before the intervention or what's causing the desaturation.

So until we start monitoring, I'm wondering whether we're going to borrow from our sleep colleagues for their definitions for how we monitor apnea and hypopnea with more of the ambulatory monitors that are coming out in the future that will be helpful, informative to then when we do the trials looking for patterns and things like that. And then eventually when we get to the FDA point and we're going the lab-based trials, maybe one of the limitations to applying it to the procedural basis was can you actually use these monitors because seemingly, we can't for some reason use the respiratory monitors or the end tidal CO2 I think

Page 106

because of mostly cost or we're just unfamiliar with it.

AUDIENCE MEMBER: We started using in a whole different way flow monitoring, measuring exhaled [inaudible -- off mic] in the air; very cheap, very easy to apply, and it's usually very good indication of the flow, [inaudible] much, much cheaper

DR. WEISS: The other question then to me, since that come up there, is the level of monitoring the same in each of these areas that we're doing? If there is not a different level, or consistent level, or a base level of monitoring, then we might be picking up different things because of our ability or our inability to pick up something that's happened.

DR. WARD: Picking respiratory, I think there are other adverse events we're interested in --

DR. WEISS: Right, but that's across the board. If we all of a sudden have to have a uniform way of picking up through our monitors, not

Page 107

just our own senses, what we're doing is, if we're not all using the same monitors, we haven't standardized it, then we're behind the eight ball there.

DR. WARD: James and then Dan.

DR. MINER: I think one of the problems we run into is when we look at devastating outcomes that occur in the community, and we go back and review for the root cause, it's usually a lack of attention, just relying on the mechanical monitor. They weren't dosing well.

If we go back and look at our clinical trials, we protocolalize [ph] our dosing very closely, we have extra people watching to collect our data, and we cause interventions that prevent most of the bad outcomes. So we do all large research trials. We don't find the bad outcomes that we see in the community.

I think that's why it's really important when we're collecting this data that we look for interventions in smaller occurrences because we extrapolate those. Well, this drug is going to

Page 108

require a lot of attention and a highly trained person to do it safely versus this drug might not because we can't even find anything when we're watching closely.

DR. WARD: Dan?

DR. SESSLER: Mark's point seems really important. We haven't discussed the minimal level of monitoring that's required for these studies, and I don't think we should get into specifying specific monitors. But it would be reasonable for us to say that in a study of sedation, you need to measure saturation and ventilation and tidal CO2, or a median tidal CO2 as a measure of ventilation. But maybe we should specify that so that there's at least a uniform dataset.

DR. WARD: I think we're going to be listening to that discussion tomorrow, but absolutely.

DR. CRAVERO: I believe even in our data analysis, we've seen that there is, even with the same monitors, variability in how well people report. When you're talking about things like

Page 109

desaturations, a sophisticated electronic medical record and monitoring system data capture gives you much more detail than if you have someone observing and just marking down when they observed a desaturation for a certain amount of time.

We've seen this when we do video analysis versus at the same time asking people to tell us about how many desaturation events, et cetera. You see different things based on video analysis versus the individual reporting. I think when you have electronic data capture, that obviously helps. However, there is artifact in there that needs to be considered as well.

There is some subtlety when you're looking at minor issues. I think the issue about what monitors you have and how you are capturing that data does make a difference in terms of how well you capture adverse events or complications as defined.

DR. WARD: Just as an aside, as a technical point, a saturation monitor is not a particularly great monitor. There's a lot of variation between

Page 110

whether it's on the finger or the ear, on the time delay that you get before the saturation is picked up.

There's a lot of technical issues about using saturation. As Albert pointed out, there's really a downstream monitor to pick up a ventilation problem. Saturation may not be a particularly good design monitor to actually do that.

DR. PANDHARIPANDE: A slightly different question, but for Anna over here. So as we're thinking about monitoring versus what are definitions of adverse event, when you think about MedDRA, when you look at regulatory requirements, when you look at industry studies versus investigator-initiated studies, so if you're going to have recommendations for clinical trials, which are done by investigators versus industry, what level of the MedDRA hierarchy would you consider reasonable?

I'll give you that example. It's coming up in clinical trials. My DSMB for my NIH-sponsored

Page 111

study went back and said, "We'd really like you to use the MedDRA classification," and then specified, "We'd really like you in your 14-center study to use the preferred term," which is 21,900 terms to try and coordinate among 12 sites when I'm the only one who has gotten a subscription.

How would you balance that? Would there be two different requirements for industry studies versus clinical trials that we are recommending investigators might do?

DR. ZHAO-WONGA: I think MedDRA does have a large number of terms, and the different levels are used at different purposes for capturing adverse events, actually at the LLT level because of the maximum specificity. The preferred term and all the other four levels are for retrieval analysis purpose.

But for sedation specific, not all 70 or 20,000 terms apply. I think that's why it's a good idea to have term knowledge like TROOPS. There are terms that are specific for sedation, and if anything falls beyond that, I would expect a very

Page 112

small percentage of adverse events would fall beyond that, then the coders can look into MedDRA to find it.

DR. PANDHARIPANDE: Just following up on that, so for example, if your patient under procedural sedation has an arrhythmia, that comes under the preferred term, which is under the 21,000 terms right now. I could classify that in the organ system and say, well, it was a cardiovascular event, which then looking across probably a senseless reporting of cardiovascular event. The arrhythmia is important, but that means I have to drill down to the 21,900 terms.

That's the balance I'm trying to say. As we recommend it for investigators, how do we try and get the balance between the two?

DR. ZHAO-WONG: That's probably going to be between investigators and the regulators in terms of how they do reporting. But for CTCAE as similar comparison, they also have a group of adverse events that are commonly seen for cancer trials. Then their guidance is these are the commonly seen

Page 113

adverse events that you use. If anything falls beyond, also possible for cancer trials, then you select in MedDRA.

DR. WARD: Dan, then Jerry; Dan first, then Jerry.

DR. SESSLER: It would be reasonable to require continuous data acquisition. It's now technically easy, and if you don't have that, you miss events, and you miss the ability to do more sophisticated analyses such as area under some threshold.

DR. WARD: Jerry and then Anna.

DR. LERMAN: One of the topics that hasn't emerged in the discussion at all is whether awareness or recall is not an adverse event in children who have sedation. Those who walk the tightrope between avoiding all these bad physiologic responses we've been discussing but keeping the child on the table run the risk of having awareness in a child. Probably more likely to occur in a painful procedure, less likely to occur in a non-stimulating sedation such as a

Page 114

radiologic procedure.

Yet it hasn't been raised as an adverse event in any capacity and would suggest that perhaps it's dismissed by this group. I guess we need to broach the subject and put it in perspective.

DR. BHATT: I could address that. In our Quebec guidelines, we actually report unpleasant recall as part of a measure of efficacy of sedation.

DR. LERMAN: As which?

DR. BHATT: As part of efficacy, so we would say -- oh, sorry, successful sedation. So we would that a procedural sedation was not successful if a child had an unpleasant recall of the procedure or a recall of the procedure.

DR. LERMAN: But it's not an adverse event?

DR. BHATT: It's not classified as an adverse event in our reporting, but it is reported as an unsuccessful sedation. I guess an unsuccessful sedation could also be seen as things did not go well, an adverse event.

Page 115

AUDIENCE MEMBER: When you hear about TROOPS later, awareness is part of that.

DR. CHAPPELL: May I make a comment from the industry perspective on that issue you just raised? We would typically report that it's lack of efficacy and it would be as an adverse event.

DR. BHATT: Can you repeat that? Sorry. Lack of efficacy.

DR. WARD: Lack of efficacy. It would get reported as lack of efficacy or reported as an adverse event.

Hannah, last question.

DR. WUNSCH: I just wanted to ask, getting at this tension around different providers having different thresholds and different abilities, has there even been an attempt to incorporate just provider anxiety or stress associated with an event as basically being an adverse event and using that almost as a way if you collect information about who the provider is, that you can start to almost adjudicate what's going on and what's causing people to get stressed out as the provider, which I

Page 116

think actually is potentially a very important part of this and could be very different certainly across providers without having to say limit certain sedation to certain types of providers.

I don't know if it's ever been discussed.

DR. WARD: I don't know much research in that area in sedation, maybe. But in oncology, there's the concept of tolerance of uncertainty and tolerance of risk. And there's actually validated measures of the provider's tolerance of uncertainty and tolerance of risk, and that has impact on the kind of conversation the oncologist has with the patient as far as the kind of chemotherapy that they're going to get.

Maybe a similar concept of tolerance of risk, maybe one of the validated forms, validated survey tools of tolerance of risk for the provider.

DR. WUNSCH: Or just even add an individual event point, does the provider feel stress by this experience.

DR. WARD: Mark?

DR. WEISS: There's a provider risk in that,

Page 117

too. Another thought might be, too, does the site of the procedure influence what we consider an adverse event as well, too? For example, sometimes we have [indiscernible] people, and the proceduralists will go up to the unit or the bedside and do a procedure, an endoscopy. Sometimes they'll say I would feel much more comfortable if they were down in the OR, we have more backup there, too.

How much does the site also influence what we're dealing with as well

DR. WARD: I'll let the panel have the last comments before we go on break.

(No response.)

DR. WARD: Let's take a break. I think we've got enough time to take our 30-minute break, so let's be back at 10:40 for the next session.

(Whereupon, at 10:11 a.m., a recess was taken.)

DR. WARD: Great. It suddenly became quiet as soon as -- speaking of duration thresholds, there seems to be a threshold value that once you

Page 118

get the noise below a certain threshold, it falls off quickly.

A little bit of change in the program. Mark is going to present TROOPS, which is the new tool that's available following on, and then we'll continue to look at common adverse events both in pediatric sedation that Joe will present and dental sedation that Ray will present.

Presentation – Mark Roback

DR. ROBACK: Thank you, Dr. Ward.

Thank you all for the opportunity to present. This is a really a work in production. This is our most recent draft, and it's tracking and reporting outcomes of procedural sedation.

Our goal is really to provide a standardized and very practical tool intended for daily use to record sedation-related adverse events, interventions performed, and outcomes. We would really like this tool to be for all procedural sedation, all types of providers, all locations outside of the operating theater, and for all age groups. Ideally, this is something that can be

Page 119

incorporated right into the electronic health records that most institutions have adopted or will adopt soon.

This is the work of the International Committee for the Advancement of Procedural Sedation. Keira and Steve presented the World Siva and ICAPS previously. Just to summarize, it's a multidisciplinary, international, independent consensus committee whose mission is advancing optimal evidence-based practice for procedural sedation and analgesia.

In this particular iteration of the committee, it's all sedation researchers from nine countries and five continents and representing, much like this group here, the breadth of providers of sedation. As we began to develop our tool, we wanted to adhere to the Institute of Medicine's Clinical Practice Guidelines We Can Trust.

Then as we started the process, we wanted to develop our definitions, and we did it through a general survey of the committee members. We based it on the previous works that have been presented,

Page 120

the World SIVA and Maala's Quebec guidelines.

Once we had the consensus, process was initiated. It was an internet-based questionnaire using nominal group technique and the Delphi method. We had sequential consensus generation with vigorous online discussion much like has been going on at this conference. We had sequential generation of our consensus of this process.

All responses from members were displayed anonymously. Revisions were based on ongoing feedback by the group. The co-chairs Steve and Keira served as moderators to guide the direction of the consensus.

The provisional tool and definitions were then submitted to outside professional societies and procedural sedation interest groups. This would be one of those, and we solicited external feedback, which was reviewed by the committee. Additional Delphi review and revision occurred, and that leads us to the tool we have today.

A summary of what we learned in the process, we really wanted the tool to be organized by organ

Page 121

systems because that's the way practitioners organize their clinical information. Outcomes other than adverse events really needed to be included as well.

Based on the first publication from the first SCEPTER meeting and some of the discussions that had been going on, clearly, there's other things that are very important, and we wanted to emphasize the patient experience, talking about comfort of the patient and recall of event. Just recall of event wasn't seen as something that was bad. Rather, an unpleasant recall of the event.

Then we also had great discussion about events and thresholds versus interventions, and we really wanted to have outcomes that were really meaningful to practitioners.

We ranked our outcomes based on severity, and as we present the primary tool today, the red would be the sentinel outcomes. These would be life threatening. They warrant immediate reporting to sedation care systems, and this should receive the highest level of peer scrutiny for continuous

Page 122

quality improvement.

Intermediate outcomes would be in yellow, and these are serious enough to endanger patients if not promptly managed or reflect suboptimal sedation quality or patient experience. These warrant timely reporting to our sedation care systems and periodic peer scrutiny.

The first part of this is the primary tool that would be used more for QI purposes and for looking at populations of patients receiving sedation, and we also wanted to recognize that we needed to have more granular data for research purposes, really building on all of the discussion that's gone on today.

These had the sentinel and intermediate outcomes as well, but we added the minor outcomes and interventions thinking that they could be important and should be studied.

This is the current draft of the tool, and you can see that we start off with the initial no adverse outcomes or events, and then if that's checked, then you're essentially done. However,

Page 123

the next part if it's yes, I think it should be emphasized that these were unplanned outcomes, and so if what you're doing is part of your everyday practice, that wouldn't be considered an adverse event or outcome.

If you checked the yes box, then you go through this table, and we have the intermediate and the severe interventions and outcomes. The first column then are our organ system, airway breathing, circulation, neurologic, and then sedation quality and patient experience.

Rob, if you could give us the online version to show you how this might work. This is how it would be in an electronic health record. That's great. Go to airway and breathing. You can see there that the definition then would come right up.

If you go over to apnea, there's also the definition. The same with pulmonary aspiration and laryngospasm.

If you can just scroll down a little bit, you can see that it does give us the -- a little bit further, please, down. There's the definition

Page 124

of the intermediate and sentinel events, and you can see the first column, circulation, neuro. Now we have sedation quality and patient experience.

Someone had mentioned we really care if the adverse event or the event results in a change in care plan. So if you could hover over hospitalization or escalation of care, this would be an important distinction to make as far as an outcome.

If you could look at the far column, Rob, maybe you can pull it over right. We thought it was very important to recognize such things as did the patient require restraint during the procedure and the sedation. This is something that we want to recognize as not optimal sedation. We also define paradoxical response, unpleasant recovery reaction or agitation, as well as important outcomes.

Next slide, please.

Then the second part of the tool would be for optional items that can be used for our phase 2 trials or for other research purposes. And then

Page 125

special clinical settings, if this is a new sedation enterprise in an institution and they wanted to more closely follow the care provided, this could be employed.

Then you can see many of these interventions that are taken from the Quebec guidelines and the World SIVA tool, tactile stimulation, airway repositioning, things that you may consider just part of your everyday practice and not being an adverse event.

In a phase 2 trial or perhaps in a sedation unit that's just getting started, maybe they want to know which drugs are leading to more of these interventions and should this change the way that we provide sedation in our specific setting.

The last slide then, we thought we would try to identify strengths and advantages of our proposed tool. Again, this is designed for widespread everyday use. It facilitates the standardization of sedation terminology, adverse events reporting, and QI monitoring.

We thought it was really important that it

Page 126

reflect, much like this group, all patients, ages, types, providers, and settings. It has to be practical and be readily incorporated into current clinical care processes if we're going to be using this as a safety surveillance tool.

However, we also recognize the importance for making it valuable for researchers as well, and that way, you could easily transition from using it as your safety surveillance, your QI project, and then it can become a research tool by adding the second portion.

Then the last piece, as we heard earlier today about MedDRA compatibility, I learned a lot about why this is important, especially as we look at our partnerships with the private sector and doing clinical phase trials.

Having MedDRA compatibility is something that we really found would be important. We were excited when Judy and Anna were able to show us that this could be adapted and made MedDRA compatible with only really minor variations.

So that is our proposed tool with the catchy

Page 127

acronym, but really our goal is to be able to track and record outcomes of what we do with sedation. Thank you.

(Applause.)

DR. WARD: We have a couple of questions now before the panel. And my question is, how is it going to be disseminated?

DR. ROBACK: How would this be disseminated? I think much like what has been done with the World SIVA tool, this could be made available as an online access if you're willing to be part of the project. And Keira and Steve could speak more to how that worked.

DR. MASON: We could decide whether or not we would accept industry sponsorship for this. It's a fairly expensive project. Just getting that tool online that I showed cost about $50,000 to have that all put online and interactive. But the nice thing is that when you do this, it's going to be all password protected, so you'll have your own way of getting into the site, and it's going to be data that you can access for yourself.

Page 128

DR. ROBACK: I think it would also be important to approach the major electronic health record providers and see how that could be incorporated into their current formats.

Yes?

DR. URMAN: Is there any plans with this tool to perhaps enable data sharing or benchmarking, looking at other people's data even if it's de-identified for research purposes, for benchmarking, something that you're planning on?

DR. MASON: As the master users, obviously, people have access to all the data.

DR. URMAN: All the data, not just your own data?

DR. MASON: We as the masters of this will have access to all the data.

DR. ROBACK: I think one of the really nice features of what they've done with the World SIVA tool is that you have this large repository, and you as an individual institution will have complete access to your own data so you can use that for your own purposes. Then if you go through the

Page 129

process, you can become part of a bigger project using these multicenter data points.

DR. MASON: I think World SIVA covers over 40 countries currently from developed and developing countries participating. You'd be very surprised. Some of the people in this room are actually actively contributing.

DR. ROBACK: Maala?

DR. BHATT: I think the tool is great. I think that it makes things very clinically relevant and easy to document, and so I think that it's a great evolution.

In reporting, if you're talking about a big multinational study, do you have a comment on how you will track denominator with this?

DR. ROBACK: I think that's a really good point, and that's one of the limitations currently of our system is that it's only numerator data. Essentially people are sending in what they've done, and those who are not sending it in, we don't know.

I think what we would do is encourage each

Page 130

institution to say is this something that we want to do and really try to get at can we get all of the data because without the denominator, it's clearly less valuable information.

DR. MASON: One thing that we did for the World SIVA tool that we considered doing for this tool is that for the World SIVA tool, if I was going to log in today and put in one of my sedation patients, it will ask me each time I log in to estimate how many sedations I do a year.

So that's the best that we can do in terms of establishing my denominator, and if it changes, then at that point, I'll change the number as I enter. But that is part of the log-in function with the provider.

You could even have -- for example, if your sedation team or your ER team wanted to be one name and one password, you want to do it as a team, you could. Then you could just estimate how many sedations you as a team do for the year.

DR. ROBACK: One of the goals of making this a practical part of your everyday workflow was just

Page 131

this purpose. So where I work at the University of Minnesota, we have eight hospitals. We're the only children's hospital. We would really like to think that as an overriding part of the University of Minnesota, that everyone participates, and that it's required, and that it's not onerous, it's just part of your workflow, and that way, we can really get that important denominator.

Yes?

DR. O'CONNOR: Just two comments. Both relate to money. The first one is that you mentioned working with the electronic health record vendors. If this were importable into the record as part of your documentation that could be used for your procedure, I think the user rate would skyrocket.

The second thing is that I'd be in favor of an outside vendor, but other data registries have used subscription fees, for example, to pay for it. I don't know if you've considered that or if that's under discussion.

DR. ROBACK: I think those are very good

Page 132

points, and quite frankly, my part of it hasn't been thinking too much about the finances. Dr. Mason is the expert.

DR. MASON: I think one of the problems about having subscription fees is that it prevents the individual user from using it, and also, people who really wouldn't -- and so a lot of them take -- like if I wanted Children's to start asking for money to pay a subscription fee, it just raises the difficulty of accessing something that we're trying to have people easily access.

But also, if we're making this a tool for all people from all countries, I think that would be a big barrier, certainly for people from developing countries. That was never our intent, and that's why we got a substantial fund from Hospira after we developed it with no hands in this at all. It was just a goodwill gesture.

DR. O'CONNOR: Those are great points. I do think if we're willing to build any coding, it would be adopted. Just making it a part of my procedure.

Page 133

DR. RIKER: Do you have a handle on how long it takes to enter the data? How intensive time-wise this is?

DR. ROBACK: That's clearly a very important part of this. I was just talking to John earlier about what they're doing with the SPS. They're done to 45 seconds on their tool. We haven't timed it, but we envision this to be less than a minute. That's been our goal.

DR. RIKER: Second point, as far as benchmarking, so we've put together an International Cardiac Arrest Registry, and when you put your data in, you've got, any time you want it, access to your own data. But you can also get access to the unidentified every data. So it doesn't tell you this is St. Joe's Hospital or this is wherever, but it gives you the group data.

I wonder if procedurally specific data for this kind of thing might be a helpful benchmark.

DR. ROBACK: I think that's a really important thing to think about. If it's de-identified, there's no reason you shouldn't be

Page 134

able to search upper endoscopy or whatever it is that you're particularly interested in.

DR. WARD: Aggregate data.

DR. ROBACK: Aggregate data, yes.

DR. WARD: Speaking of data, Joe's going to give us some real data related to adverse events in pediatric sedation.

Presentation – Joseph Cravero

DR. CRAVERO: This is a great discussion for me. I would just say we've been talking about common and important adverse events within the groups that I've been working with for at least the last 15 to 20 years, and the discussions have gone quite a bit like what we've done today.

What I'm going to try to do is just present generally the data that we've had, and I'm not trying to orient the discussion other than to say this is what it is. I would encourage us to think maybe about how this applies to clinical trials specifically because I think there's a real issue with talking about our research and our quality data versus what we want to know from a clinical

Page 135

trial. As we're trying to start a new clinical trial at our place, I see a real difference between those two things.

Just for a quick comment, I do think pediatric clinical trials are slightly different. We understand that essentially all infants, toddlers, young children, older children with developmental delay require sedation in order to contain their emotional and other issues around the procedures that we do. We have to sedate kids for non-painful procedures that you just don't do for adults generally.

We really do have a slightly different patient cohort that we're dealing with, and a lot of these patients have very little pathology that would impact on their sedation, whereas I think again the majority of your adult patients have a lot more comorbid issues and you're sedating probably less for the MRI scans and more for more invasive types of things like your upper GIs and other stuff.

Just thinking about the fact that we are

Page 136

sedating kids in remote locations and that the demands are slightly different for our kids than adults.

I'm sorry. I have talked about this so many different times in different groups. Many of you have seen me talk about exactly what I'm going to talk about here for a second, but they asked me to give a talk. This is what I got.

I got to say what I usually say, which is I would encourage people that the sedation literature on clinical trials generally reports events that range widely. They do record a lot of physiological disruption, including O2 desaturation, which is the most common thing that is reported, and as we've already discussed, I'm just very uncertain about what it means.

They do talk about airway interventions, and we do get reports on how many kids require positive-pressure ventilation, et cetera, which I would say is important and interesting. Maybe the problem becomes when we start using taxonomy like complication or adverse event.

Page 137

Within the PSRC, we've really tried to get away from that because it's so loaded that we now go toward what do we just want to know, and we now record things like what interventions were required during these sedations without any of the judgmental implications of using the idea of complication or adverse event.

Whether an anesthesiologist is readjusting the airway or an emergency medicine person is doing that or whatever, rather than getting into is this a complication or not, we're just talking about what needed to be done in order to get MRI scans done with propofol -- that's what we want to know -- or dexmedetomidine, or whatever, and then you can make your own judgment about how important those reports are.

Just give you a couple of examples, I don't use these as bad or good clinical trials, just this is the kind of thing we see in pediatric sedation. This was a report of propofol in the pediatric intensive care unit. It actually was a comparison or propofol versus ketamine for rather deep

Page 138

sedation -- I would call it general anesthesia -- in the intensive care unit with sedation provided by intensivists.

In this particular case, the report was that 12 out of 58 patients required airway manipulation, 10 required positive-pressure ventilation, 3 out of 47 of the ketamine group required positive-pressure ventilation, which is a little different for me. I think that's a fairly high rate for ketamine. 1 needed to be intubated because of what was described as "difficult ventilation." It wasn't really described more than that.

Again, I'm not trying to make judgments here. I'm just telling you this is the kind of thing that is reported in clinical trials concerning pediatric sedation.

Another trial, again, in part of this trial, they actually recorded all the different types of interventions that were required. And I'm sorry for those of you who can't read this, but there's things like airway repositioning, apnea that required bag-valve mask, intubation, et cetera. So

Page 139

this is very typical of pediatric clinical trial reporting.

I would offer a question as to whether or not you consider these things adverse events. 10.6 percent of the ketamine group experienced what was thought to be discomfort during the procedure. Again, that may be considered more efficacy than adverse event. I think there's a little bit of a gray area there as to what's efficacy and adverse event reporting.

23-minute recovery time for propofol, 50-minute recovery time for ketamine. Again, our thinking in the pediatric sedation research consortium is very extended recoveries do represent an adverse outcome. You can argue whether that's actually true or not or is that really some measure of efficacy, but we do think that it's important to think about sedation regimens that require hours and hours of recovery. Is it a significant thing that we need to think about in a clinical trial? Vomiting, et cetera, similarly.

Another clinical trial looked at -- this was

Page 140

an observational report -- of propofol used for emergency medicine provided elective sedation for hematology oncology procedures in pediatric patients. In this case, it was propofol procedural sedation.

It was a prospective evaluation of 393 sedations. They reported 5 percent of their patients had hypoxia as less than 90 percent during the procedure, 3 percent required airway manipulation, meaning jaw thrust or head tilt, and 1 percent required positive-pressure ventilation.

I think this is very typical of what we see with propofol and as a clinical trial outcome in children. Whether or not you consider any of these really complications or adverse events, I think we could again go on probably all day.

The conclusion, as they almost always are in these clinical trials, is that drug X is safe and effective for procedure Y. In groups of 393, I would offer that that is a fairly small group to try to conclude that a given technique as a clinical trial can be generalized to the entire

Page 141

population of children undergoing procedures.
At Dartmouth -- and I think I've talked to a lot of you here about our work there -- we really tried to take a more human factors approach and look at a very detailed analysis of the way sedation is given. We came up with a way of thinking about the goal, which is to get a child through a procedure such as an LP, going from your starting point to a similar ending point with the same level of consciousness and health.
During the course of that procedure, you're going to have side effects due to pain. You're going to treat that with either morphine or other types of sedatives, and you're going to be getting yourself into side effects and/or adverse events related to undersedation or side effects and adverse events related to over-sedation.
So as part of this, we came up with this Dartmouth Operative Condition Scale, which I think I presented the last time we met, which judges the conditions of the patient during a procedure based on pain, stress, movement, consciousness, and side

Page 142

effects from the sedation, which we defined as saturations -- I'm trying to read this. I'm getting old -- and respiratory pauses, and then low blood pressure. So we tried to incorporate all aspects of the child's status related to the sedation itself in one scale.
We have just submitted to pediatrics -- and I think it's conditionally accepted -- a new scale that will be the procedural sedation scale for children or PS3, which has six levels from zero up to 5, which considers the state of the child. Either they are wildly out of control, experiencing problems from undersedation, to a state where they are out of control, experiencing too much sedation and physiological disturbance in spite of intervention as a zero. So there you are providing positive-pressure ventilation, but the sats are still abnormal. We grade it from low to high.
Again, during the DOCS validation, we tried to define three zones with when you add up our scores, you can either have a high score that's associated with side effects from the procedure or

Page 143

undersedation and very low scores that indicate side effects from the sedation itself. We've published this work in A&A.
I would just offer you that when you do this and look in a very detailed manner at the scores that you get over time and overlay the time of the procedure, you get a better idea of how you were meeting the demands of the procedure with your sedation than you do when you just have intermittent reporting.
This goes a little bit to what we were talking about before. We looked at this scale, published a study looking at the scale over 110 different procedures, assigning DOCS score every minute to their procedures. And we found that the failure to achieve sedation was about 5 percent. It was 8 percent when you didn't have expert providers, and zero percent with expert providers, defined as those people that provide sedation as part of their professional work as a team, so basically sedation service providers.
We found that there was huge differences in

Page 144

the time from beginning of the sedation to the time of the beginning of the procedure, and that can vary on the effectiveness of the sedation activity. We almost consider this an adverse outcome when you're waiting that long to start a procedure, but it probably is better classified as effectiveness.
We did classify over-sedation events and undersedation events, and I guess again, you could call this all under the rubric of effectiveness. But we think there really are problems or there are complications associated with undersedation.
Again, going to what we were talking about before, when we look in detail at these tapes, we found issues related to low sats for significant periods of time that were not recorded in the record of the patient that had been sedated. We found kids that were not fully recovered when they were discharged, yet the recording from our nurses or the nurses that were involved was such that they indicated the patient was ready for discharge.
I think as we were talking before, the definition of states, the electronic capture of

Page 145

data is going to be really important in clinical trials to actually know that you're talking apples to apples, oranges and oranges.

Under-sedated states, again, we could call this efficacy. We were calling it adverse event at the time, but we found that very commonly. I think in pediatric sedation, the undersedation of patients is actually a much bigger problem than the over-sedation of patients if you look at outcomes generally.

Just very briefly, I think many of you have heard me talk about Pediatric Sedation Research Consortium. It is a consortium of specialists across the country. We have 48 institutions involved, about 20 percent anesthesiologists, 33 percent intensivists, about 30 percent emergency medicine, and 70 percent other specialists, largely hospitalists now that are providing us information.

We do collect a lot of different data elements in this project: patient factors, procedure factors, sedation technique, care providers, observed care, and in specific germane

Page 146

to this lecture, complications associated with each sedation encounter. The complications that we collect include apnea -- and I'm sorry. In blue here are the definitions that we've had. And I guess I'm not sure about the World SIVA effort, but we found that being very specific when you are talking about adverse outcomes is incredibly important because if there's any way to misinterpret something, people will. I'm not kidding. Even death.

(Laughter.)

DR. CRAVERO: It's amazing the variability. "Well, what did you mean by dead?"

"I don't know. I think it's fairly clear."

It's amazing. So over time, we've had to really be very specific. And what we did ultimately was our interface for our data collection tool has a bunch of click boxes that you click for the primary problem that you are taking care of, the coexisting medical problems of the child.

As was just indicated in the last lecture,

Page 147

this tool now -- by somebody who knows how to use it and there's a lot of skill that comes with doing it a lot of times. But you can fill out this tool for the average pediatric sedation in less than a minute. But if you hover over any one of these things -- and I'm sorry, I don't have it active on mine. But if you hover over any one of these, it will give you the definition so that there is as little confusion as possible.

I would say in clinical trials, this would be absolutely critical that people understand when you click laryngospasm, what exactly do you mean, and when you say hypoxia, what exactly do you mean. It just is a morass if you don't have that well laid out.

Just to deliver on the promise of data, we've published about 15 papers out of this effort now. The first one was as exemplary as any of them, talking about the incidence and nature of adverse events during pediatric sedation. Data collection at that time was 30,000. Now we have almost half a million encounters in our database.

Page 148

At the time out of 30,000, we had zero deaths, 1 cardiac arrest, 1 aspiration, 24 stridor and laryngospasms, 21 unplanned admissions. So we considered these indicators of major issues, and it's not very ambiguous that there was a significant problem when one of these things happened. We were able to say that happens about 1 out of 1500 sedations in kids.

There were some less serious adverse events. Stridor and laryngospasm, wheezing or apnea that interrupted the procedure was 1 in 400. 267 vomiting, secretions, or desaturation episodes that interrupted the procedure, meaning somebody had to stop what was going on and address these issues, about 1 in 100 sedations.

We did a similar kind of study looking at the incidence and nature of adverse events using propofol from a large group of our reporting institutions. In this case, it was over 50,000 encounters, and again, just to give you a flavor, 6 emergency anesthesia consults, 29 emergent intubations, 83 oral airway insertions, 192

Page 149

positive-pressure, bag-mask ventilation.
To the specific issue of clinical trial -- and I'm sorry, John Berkenbosch over here, suffered through discussions that we've had for hours about this. We've really gotten away from trying to determine was this really an adverse event or wasn't it, was it expected, was it unexpected largely because that language becomes so difficult to be precise about that we have ultimately decided we're just collecting this, and that's what we're going to report. Then people can make their own decision about whether or not this represents an adverse problem or not.
I would argue for me in my job, the fact that I had to insert an oral airway, not a big deal, but that could be something that you want to know in a phase 2 trial about the use of propofol for sedation of children undergoing MRI scans.
That's what we've ultimately come down on, and while the title of this talk was supposed to be complications, I'm almost like I don't even want to use that word because people start to get very

Page 150

uncomfortable with it. Clearly, there's a few things we can say are complications, but a lot of things that are ambiguous when you try to use that language.
I'm going to skip through this because it's sort of repetitive, but for Dr. Twersky, just so you know --
DR. TWERSKY: I know you don't like --
DR. CRAVERO: You're getting it. I'm sorry.
We did consider inadequate sedation a problem. This is rate per 10,000, and this is the absolute number of problems that we recorded in this 50,000 patient group. You can see we have things like inadequate sedation, airway obstruction, allergic reaction, apnea as defined as greater than 20 seconds during the course of a procedure, agitation at the end of a procedure. There's a bunch of them here.
Interesting, I guess I'd just point out, some things we did not include initially were things like IV complications. We did not have this as one of our elements to collect initially, but so

Page 151

many people were writing in that they had problems with IVs during the procedure, that they infiltrated or they couldn't get an IV started, et cetera, that we then started to include IV complications as one of the issues.
We also had to include, say, secretions. We did not have secretions down as a problem during your sedation, but people started writing this in so often that we ended up having to include it. We defined it as secretions that required you to interrupt the procedure and suction in order to maintain stability and easy respiration within the patient.
I would say to you there are things that come up, and again, is that a complication? I don't know, but you probably want to know how often that occurs with drug A versus drug B or drug combination C. So I would say I'm not calling it a complication, but I think it might be something you want to know.
We did collect cardiac arrest data, and I'd just say it's interesting. Again, cardiac arrest,

Page 152

you'd think that's clear, but then there's all these things that happen like significant bradycardia that was profound. Maybe there was asystole, maybe not, but CPR and epi was given, therefore, we considered it a cardiac resuscitation.
Maybe not everybody would have gone to this level with the heart rate of 25 or 30; maybe they would. I'm not sure, but you have to be careful about exactly what you call these things.
We had another 16-year-old athletic male who was having a colonoscopy, got very bradycardic, and was considered to have asystole for 30 seconds. CPR and atropine was given, and the kid was back to baseline in 30 minutes. So very recoverable things, but clearly, here's a major problem that we want to know about.
We also collected unplanned airway interventions, which we have morphed now into just airway interventions because we considered this over time. The unplanned part of this just became too hard to know. So now in the current

Page 153

publications, we just talk about what kind of airway interventions were required.

We have a fairly recent paper looking at major adverse events in relation to nil per os. We published this in Anesthesiology just last year. In this case, we looked at 120,000 patients undergoing sedation, and then we looked at whether or not they were meeting NPO criteria or not.

For the purposes of this talk, I just want to say we had a fairly specific definition of aspiration, and that is, you saw contents coming out of the mouth during the sedation. And then after the procedure, you had a change in status that was significant, requiring oxygen, requiring admission and/or x-ray evidence of a problem with respiration that was not anticipated.

We also were recording major adverse events. I would just say to you we decided from a observational data collection that we're going to stay with major complications which we define as cardiac arrest, aspiration, unplanned admission because it gets very fuzzy when you get into the

Page 154

very minor problems given all the things that have already been discussed here today.

But when you look at aspiration and major complications, I can just tell you that we were not able to find a correlation between the patients that were, in fact, meeting NPO criteria and those that weren't. Now, we could get into a whole discussion on this. There's all kinds of possible confounders here. We did do multivariable logistical regression to try to get rid of them, but as far as we could tell, within this group, we were not able to determine a direct relationship.

But I think more to the point of this talk, we have tried to be very careful about how we define adverse events and the reporting of things mostly in terms of major adverse events that are easily agreed upon.

I'm going to summarize here that I do feel like pediatric sedation adverse event is slightly different because of the nature of our patients and the nature of our practice. If you're looking for heart attacks, you're not going to find them

Page 155

really. If you're looking for more minor things, you're going to find a lot of them.

How we consider them I think is an interesting conversation to have in the context of clinical trials and possibly getting away from necessarily calling them complications. There are many reported adverse events. I would suggest many of them have little or no meaning from the outcomes perspective, as we've already said.

There are many minor complications, and I do think we need standardized definitions that include some physiology with intervention. I obviously agree with a lot of the conversation that's gone on so far that's morning.

I guess I'll take questions when we have the panel.

DR. WARD: You can take a couple questions now.

DR. CRAVERO: Anybody have any questions?

I am obviously very steeped in this stuff, and we've talked about this so much. I'm interested in the conversation we've had so far

Page 156

this morning.

I do think there's a real difference, as you pointed out earlier, between what we collect and what we're probably interested in from the quality improvement standpoint or from an observational database standpoint and what I want to know as a clinical trialist when I'm comparing one technique or one drug to another. I do think as we go along today or tomorrow, that kind of perspective needs to be considered.

DR. WARD: Apnea was one of the more common or cessation of breathing?

DR. CRAVERO: Right.

DR. WARD: Was there any definition of how that's measured? That's actually without direct physiological measurement.

DR. CRAVERO: Yes. We have it under -- if you hover it, I think, but it's lack of air movement for 20 seconds or greater, is our definition of apnea. And it does not necessarily imply central apnea or obstructive apnea, which is obviously a problem.

Page 157

But when you're collecting data from 40 institutions, which you're probably not going to do with a clinical trial, but my personal feeling is it is very hard to get good data from a large number of institutions. So you have to be very cautious about the conclusions you make based on data that comes from a real large variety. So we did not try to get into the subtlety of was it central or was it obstructive, just did you observe lack of air movement by either your direct observation or end tidal CO2 for 20 seconds or greater.

DR. WARD: Dan?

DR. SESSLER: This reminds me a bit of airway device evaluations, that the reason that you're interested in a novel type of airway device, let's say a video laryngoscope, is the hope that it will save you when you get into a can't intubate/can't ventilate situation, that is, when your patient is trying to die, you hope that you can reach for some device, and it will save you.

The trouble is that these events are very

Page 158

rare. Great that they're rare, but it makes it really hard to study them. So there has never been a study of any airway device that remotely addresses what we're really interested in.

Instead what we have is lots and lots of studies -- I'll admit to having done some of them -- where you have intermediate outcomes such as time to intubation. The trouble is that that's not really relevant. It's not really interesting, but there are hundreds of articles that evaluate different airway devices with time to intubation as the primary outcome.

I'm a little concerned that we have the same potential dynamic here, where you look at something like desaturation, which is only tenuously related to the things we really care about, which are serious complications, patients trying to die, and your data are very encouraging because they don't seem to be dying, but it tells you it's going to be very hard to study.

But we need to be careful that just because the outcome of interest is difficult to study, that

Page 159

we don't just slide into some outcome that's actually not interesting and perhaps unrelated to what we are interested in.

DR. CRAVERO: A couple of obviously great points, Dan. I think there's a couple things from our stuff, which is we preface any report that we make that we are talking about high performance sedation services functioning in primarily children's hospitals but also large community hospitals, very few small community hospitals. I think people need to take that data for what it is. It does not imply that this indicates what the data would be if you looked at the entire country. I think it's a good point.

Secondly, I think, again, you get to the issue of what do people need and want to know in a clinical trial. Our point in the Sedation Research Consortium has been more what do people need to be able to do and what do they need to understand about sedation practice in children.

If you're providing propofol at the level of 200 to 250 mics per kilo per minute to children of

Page 160

a given age, what do you need to be able to do, and what can you expect to see when you're doing that? So you need to be able to recognize apnea. You need to be able to open an airway, and rarely but too, too rarely, you need to be able to provide position pressure ventilation.

My thought is when you talk about clinical trials, you probably want to know -- although I would agree with you, that still doesn't really tell me about the outcomes I'm most interested in, like did anybody need to be admitted that shouldn't have been admitted or did anybody die or whatever. But it probably is important information when you're trying to compare one drug to another.

So if I look at dexmedetomidine for a group of patients versus propofol, I might want to know how many times do you have to intervene with dexmedetomidine versus how many times you have to intervene with propofol when I want to think about how I'm going to use that drug clinically, who's going to use it, how are we going to use it, et cetera, even though I would completely agree

Page 161

with you, it still doesn't tell me which drug is safer. And I think we need to be cautious about using that kind of language. I totally agree.

DR. SESSLER: It gets back to who's doing the procedure, who's doing the sedation, and the context of the intervention. So, for example, something as simple as providing oxygen is a disaster if you have to stop an MRI to put oxygen on. In another context, providing positive-pressure ventilation, if it's an anesthesiologist doing that, it's trivial. It's part of the job.

DR. CRAVERO: I would just say again, for us, our whole effort, since we're a multispecialty group, is to try to, as much as possible, get away from the contextual part of it when we report stuff, and just say this is what happens. You can make the decision as to how important those things are or how worrisome they are, but we're going to tell you when you use this drug for this type of intervention, this is what happens, at least in the group that we see.

Page 162

DR. WARD: Rick?

DR. RIKER: I just want to jump off on Dan's comment. So for rare events, I think maybe we need to overtly say that a randomized trial is not the gold standard or shouldn't be the gold standard. For something that happens 1 or 2 percent of the time to look at 99 percent of your data and not find that event, maybe what we think of as a lower quality type of study, a registry, a cohort based on a specific outcome, and drilling down in that situation may be better.

We did a propofol infusion syndrome study in the ICU and wasted most of our effort on the wrong patients. I think a careful consideration of what the right research design would be for these uncommon events is worth discussion.

DR. CRAVERO: I think you're right. For pediatric section, that may not really be from a clinical trial perspective, something you're going to even be able to do. You're going to be able to look at some of these outcomes because they come with a frequency that you can actually look at and

Page 163

consider in populations that are possible to do with clinical trials.

You can try to get a frequency for the really rare events by using the kinds of things you're talking about. But again, I would personally like us to think about, if we're going to be discussing clinical trials, what are the reporting requirements and what do we think is useful in that context, which I would say if you're looking for neurological injury due to sedation accidents, the clinical trial is not going to do that for you. You're going to have to understand that you're not going to get that out of this.

DR. WARD: Speaking of areas, another area we maybe as a group doesn't necessarily think much about as being done a lot is the dental sedation, certainly an area in which those of us who have had root canals and maybe a long time ago your wisdom teeth taken out, have had to put up with dental level sedation.

Ray has done a lot of work with that, and we want him to talk a little bit about adverse events

Page 164

in dental sedation.

Presentation – Raymond Dionne

DR. DIONNE: Thanks for having me. I retired from the NIH about three years ago. When I look back at my career, I say, gee, I had a successful career, but I accomplished almost nothing. So I went down to East Carolina University and was kind of wasting away going to seed. And for the misfortune of society but for my good fortune, opiate overdoses and deaths associated with sedation have become noteworthy recently. So it's given me a renewed career. I'm sort of a born again crusader.

(Laughter.)

DR. DIONNE: I have one bad slide here. I'm going to see if I can get it all there. This is the problem with not having anybody to do my work for me anymore. I actually have to do these things.

You might ask yourself, why is even sedation needed for dentistry because if you're of a certain age group and a certain SES status, it's not a big

Page 165

deal anymore. However, the fear and anxiety about dentistry is still very prevalent in the population. It seems to originate in childhood for good reason. Those needles that you get seem to be about 6 feet tall, and they hurt quite a bit. Then all the stuff that follows after that can be very unpleasant as well.

It does appear to lead to the avoidance of dental care, and it's remained stable over the past 50 years despite all the progress we've made with preventive techniques and improved restorative techniques. It also seems to be that if people do have high dental anxiety, which is about 15 or 20 percent, they'll go to the dentist less frequently.

Now that we have some fairly good association data that suggests oral diseases are related to possibly cardiovascular disease and diabetes, there might be greater implications that just a little disfigurement and early onset of a denture or something like that. There may be more going on.

About 21 percent of patients in the survey

Page 166

we did a number of years ago said they would definitely go more often if they could get some kind of drug that would make it a little bit easier for them to tolerate.

One of the problems I've had since I became a dental educator in my new role is at first I thought I was going to shape their young minds. I got over that delusion about four lectures in. So now what I try to do is scare them at the beginning and hope they'll pay attention at least to the 15-minute mark or so.

(Laughter.)

DR. DIONNE: I did this little thing right before I was going to give a talk on anesthesia and sedation to the dental students, and I was startled to have this stuff pop up. This was like the first two pages of a Google search, and it talked about children being killed. Apparently, if you die undergoing a procedure in a hospital, that's a death. When it happens in a dental office, it's murder as portrayed by the literature, or by the public thing.

Page 167

One thing that caught my eye also was this handling of cases about questions about the state review process, which is down there. And I've recently discovered there's two states where there's a big controversy brewing where they think there's between 40 to 50 deaths in the last 5 or 7 years, depending on what report, that have been swept under the carpet in both of these states. And investigative reporting has suggested there's an issue there, but no one's been able to pry through the liability insurance data that's always closed and forgotten apparently, and the state data.

It may be that the things that we do see in the public domain only represent the tip of an iceberg that may be a lot bigger than I ever would have expected.

Then there's another thing that's implied by this is that there's been a growth of people who use sedation as an aggressive part of their marketing process to try to get people to come in, and this has become probably resulting in too many

Page 168

people being exposed to these procedures. Worse yet, the procedures that they're pushing are far removed from evidence based.

This is some old data that came from Charles Cote, who I always think of as a friend of dental anesthesia and sedation, but he always seems to publish the data that makes us look a little bad.

(Laughter.)

DR. DIONNE: This was some stuff he published related to case reports he could find in the FDA database that was available in the USP and then reviewed the literature. Here is the way he -- there was 95 cases, and because there were many things that contribute to any particular situation, he had far greater numbers.

What I tried to do for purposes of teaching is point out that these two leading categories are really drug and dose related. Our profession seems to be obsessed with training people to do resuscitation better, not to address possible preventive procedures associated with that.

Then there are procedures and methodology-

Page 169

related things, and then finally, there are training-related things. Very often, the question about the personnel arises because there is still a standard within the dental community that if you have a dental assistant who has taken one week of training, they can be the surrogate for the anesthetist/anesthesiologist as long as the person who's the captain of the ship is trained up to that level of anesthesia/sedation. But of course, the captain of the ship is over here doing a procedure and monitoring, and even the drug administration is often done without full supervision. That's a little bit of a problem as well.

When you look, this is old data, you can't deny that dentistry seems to be the leading perpetrator here. About a third of the deaths were associated with dental procedures, so more likely to have serious morbidity and mortality. Like I said, my friend Dr. Cote points that out. Actually, he wouldn't remember me if he walked up and saw me, but I at least like to throw his name out like that.

Page 170

All right. So what I did recently, I got myself in front of a train that was going down the legislative process at the American Dental Association. They were going to put to a resolution that was going to regulate or I thought over-regulate the safest form of sedation, which is nothing more than having people swallow a benzodiazepine but ignore everything else.

I looked at the same strategy. Could I scare the people by doing a little literature search? I just did a literature search on deaths in the dental office, and I was startled to get 600,000 hits. Now you know the ratio of meaningful stuff and garbage is pretty high, but I started plowing through this. And only in the first 500 reports I reviewed -- and then I had a second person go through and do the same thing, and we came up with some agreement on it -- was that there were 42 deaths. I'm not going to extrapolate that there's 20,000 based on this kind of mathematics, but it's worth digging in. If epidemiology follows big footprints, this may be a hint that something's

Page 171

going on.

If you looked at the age range, it was a lot of kids, mostly less than 2, up to 89 years, so 28 out of the 39 where the age was reported were children. As far as the modality used, general anesthesia was associated with 20 of the deaths; IV sedation, 13; oral, 5; and that was the modality that people were going after. "We've got to stop something. We've got to stop this scourge of oral sedation that's causing all these problems." Only 2 of these were actually associated with triazolam, which was the drug that was being implicated.

Nitrous oxide, which is hard to imagine unless you have a plumbing problem, but if you give enough local anesthetic with it, apparently you get problems, and then 2 were not reported.

So then I looked at the practitioners, and again, it's hard using reports in the public domain, but for every one of these, there's usually 20 reports. If you read through all of them, you can get some idea of what's going on.

Oral surgeons, who are probably the greatest

Page 172

users of general anesthesia in the dental profession, although I think they only represent about 6 percent of the total number of dentists, were associated with 10 deaths. Pediatric dentists, also frequent users of sedation, 7. But what kind of surprised me is general dentists were implicated in 13 of these deaths. Then others were either not reported, but we're still digging on that.

My bias, of course, is it has a lot to do with the drug. So I looked as closely as I could at the drugs. A lot of benzodiazepines reported. Diazepine given alone at a reasonable dose usually doesn't cause problems, but then they're almost always associated with an opiate. And as the FDA has pointed out in their recent warnings about the combination of opiates and benzodiazepines, it's a different picture when you put the two together.

General anesthetics were being used. Chloral hydrates, which everybody tells me no one uses anymore, somehow or other is still causing reports of pediatric deaths.

Page 173

Then as far as the combinations, benzodiazepine and opiate was almost always what was reported, but then you had the multiple drug combinations. Now, this is kind of an improvement because I did a survey about 25 years ago, and it wasn't uncommon to find people reporting that they were using five or six drugs on a routine basis, so at least things have done in a little bit better direction.

The most problematic thing in this area, which I don't know we design a safety endpoint for this, is the single operator anesthetist. This is still considered to be a professional entitlement for some people. You have the person who is doing everything himself, or if you're doing general anesthesia, then you're obligated to have an assistant, which as Cote did report one time, he says, "That's like having a high school dropout with one week of training" was the way he characterized and saying that's equivalent to medical anesthesiologist.

Then there were a surprising number that had

Page 174

a separate anesthetist/anesthesiologist, either an RN, MD, or a DDS, and there are a small sliver of dentists who get two years of anesthesia training and are supposed to provide that service as separate from the operator.

Causes of death were almost always either respiratory depression or cardiac arrest secondary to respiratory depression so that was pretty common. If, in fact, there's any credibility to this kind of crude way of doing things, something may be going on that suggests there's a lot more morbidity and mortality associated with sedation than I ever would have imagined, and that it's kind of being swept under the carpet right now by those people who have the professional benefit by being able to promote this as part of their repertoire.

What else might be going on here? This word "sedation" always bothered me because I always thought we were trying to produce anxiety reduction. Even for a while in the dental profession, they used to call it anxiolysis, which was that category at the low end of the dose

Page 175

response, somewhere above nitrous oxide, somewhere where oral sedation might fit in, but before you get into parenteral and whatever. At least there was recognition then.

This was a monster study we did, had 1,000 patients in a prospective, five sites. Got the government to spend a lot of money on it, and I thought once this study was published, well, my work was done. When I came back years later and looked again, nothing had changed on the basis of what I had promoted or published.

This was a measure of efficacy. It wasn't quite anxiety reduction, but it was a global measure. What you could see is if you had a placebo but always with local anesthesia, which is an important distinction from a lot of the things that are done in the medical world -- we almost always have to give effective local anesthesia to perform our procedures.

If you just gave them midazolam, titrate it to a clinical endpoint that would be considered sort of light sedation, you got a rating there. If

Page 176

you kept pushing midazolam every time the person wiggled, you got a little bit of an improvement. If you gave midazolam plus an opiate fentanyl, seemingly in the same range of efficacy as judged by the patient, and then finally if you produced deep sedation with a combination of midazolam, fentanyl, and methohexital, you got -- these were all within the same range of efficacy as judged by the patient.

However, when you looked at the observer's rating, and we had the person doing the procedure as well as a separate person who was just there to observe the patient, clearly, they thought the more CNS depression you produced, the better off the patient was or the more cooperative they were or the better sedation.

I always get a little nervous when we talk about what the operator wants versus what the patient wants when I see this because there's no increase in benefit from the patient's point of view, yet there is the potential risk associated with giving two and three drug combinations, and

Page 177

that seems to be supported by that case report data I just showed you.

The other side of the equation, is there some safety consideration? Well, these are the people that got the placebo or the two midazolam regimens, and you don't have to be able to see too far. Even with my 69.9 years of vision, I can tell that that looks a lot different from that. And these are the people who are having 100 percent oxygen supplemental, by the way.

So respiration rate went down. Oxygen saturation went down, and expired CO2 went up. So it suggests then that the potential risk from these drugs in combinations that depress respiration are not providing any benefit to the patient if they're admitted, just an anxiolytic drug and given effective local anesthesia.

All right. So you'll say, well, that doesn't make any sense because we know people are having pain, the big joke about the endo, the root canal procedures, the extractions and things like that. And granted, if you're having it done with

Page 178

inadequate local anesthesia, that's a very painful process. But if you do adequate local anesthesia -- and it's not that hard to achieve in the mouth, 95 percent success rate on the first shot, by the second shot, it goes up to 99, and if you're still missing, the third one is always magical.

I one time was having a problem with a patient, and I went to my colleague. I said, "Gee, Dave, I don't know what to do. I've given this mandibular block twice. Should I give him another one?" He said, "Ray, there's two forms of anesthesia, numb and not numb. You got not numb. Give him another shot."

(Laughter.)

DR. DIONNE: So I overcame my anxiety as a pharmacologist and give the third shot. Magical. Sooner or later, you're going to find it or it's going to move around enough that you get it.

We did a study, and we were using at the time a scale that Rick Gracely had developed, where he had demonstrated that he could separate out

Page 179

sensory intensity, how much they felt, how bad it bothered them, which he called unpleasantness, and then their overall pain report.

So if you gave people local anesthesia and a placebo and took out two teeth, you got that kind of a pain report, not very high. This scale is hard to interpret, but this was in the ballpark of slight pain or slight sensory intensity.

If you gave them diazepam and then took out two more -- you couldn't do a crossover here obviously because the diazepam wouldn't go away in that short period of time -- or if you gave them fentanyl, there didn't seem to be any difference in the sensory intensity.

However, if you asked them look at the unpleasantness of the sensations, diazepam was clearly having an effect, and this wasn't a recall thing because we were asking them 30 seconds after the procedure was over to give us these ratings. Fentanyl did nothing because of course, there was no clinical pain to speak of for it to relieve, so it didn't do much.

Page 180

Then if you asked the patient overall what was your pain during the procedure, diazepam actually, using a category scale, took it down to no pain reported, whereas fentanyl just had a marginal non-significant effect.

This would suggest then in the context of dental sedation with local anesthesia that just the opiate alone doesn't give you much additional benefit, but it does have the risk of respiratory depression.

Having looked at this literature for a long, long time, I tried to parse it out into what were the factors that were determinants of safety, which would be presumably the things we would try to come up with as safety endpoints when we were trying to design studies.

While no one believes me in the dental world and I find this incomprehensible, I think the drugs have something to do with safety.

(Laughter.)

DR. DIONNE: PhD in pharmacology, you didn't waste your time, Dr. Dewey. I figured it out. I'm

Page 181

carrying that knowledge forward.
The doses obviously make a difference, root of administration, parenteral versus oral. People worry, oh, sure, if you load enough down there, it's going to give so much sedation that they'll be just as obtunded as they would be if they got parenteral, but even then you got advantages of slow uptake and the beginning of elimination.
The rate of administration, which was proved by a dentist in North Carolina few years ago, where he gave 10 milligrams of midazolam as an IV bolus, and then he pulled out his butterfly and started his procedure, and quickly found he was working on a dead patient. While people still question me about that, I think that one's pretty obvious.
Then the combinations of the drugs, one of my colleagues years ago looked at single drug, two drugs, and three drugs given for pediatric sedation. He referred to the three-drug combination as the "kid killer cocktail" because it was always the one that was associated with the significant morbidity and mortality, whereas the

Page 182

single doses of those or even two combined usually didn't cause the big bad outcome.
Patient selection, preoperative value makes a difference, obviously monitoring, premature discharge, all these things. But for the purposes of trying to focus in, I consider everything that has an asterisk a preventive factor that maybe we could use when we're trying to teach people or ideally set some guidelines for how people do things that we might take those into consideration.
If you take all that together, I have my little pyramid here, which I always try to simplify everything down to that level. I work on the theory that the students only remember at most two or three things that I say over the course of 50 minutes, so I try to make it real simple.
I try to point out that this is the foundational things. And even in the most skilled hands with the best training and experience, you can't always overcome that finite incidence of problems that happen.
There was a classic example in the dental

Page 183

literature years ago of a guy who cooked up his own little technique, went around the world telling everybody how great it was. He did it like 10,000 times, and the next case is when he had his serious adverse event. And he had no idea how to treat it because he'd never had one, and he wasn't trained to that level. So he had a death, and the next thing you know, I'm reading about his coroner's inquiry that gets published in the British medical journal or something like that.
So all that good 10,000 cases safely, still if there's something that's inherently dangerous about the method, it manifests eventually. And all this stuff also, clinical judgment makes a huge difference, and I don't know how we can come up with a risk factor for that. Can't be giving the guys MMPIs ahead of time.
This is where the balance is, I think, between safety and therapeutic efficacy. I think for purposes of moving forward, it'd be nice to recognize that opiates do produce obviously a dose-related decrease in respiration. General

Page 184

anesthetics obviously do. In the right hands, in the right context, or with the right risk, explain to the patient because very often, this is done with the assertion that these techniques are safe because I've used them my whole career.
There was one case report I read of an 80-year-old oral surgeon who had a death, and his defense was, "I've been doing it this way for 50 years. I know it's safe and effective." Well, that particular day, it didn't work out, and he had a young kid.
That's the other thing that's a little discouraging. You're used to seeing people that have got medical indications for this. You read these case reports, and it's just one picture of a young kid after another with handwringing by the parents and the journalism making a big deal out of it. It gets a little depressing.
Local anesthetics, it's hard to cause respiratory depression with that. In looking at all the case reports, I could only find one that seemed to be a very high dose of mepivacaine. The

Page 185

additive effects are obvious, but people tend to sort of ignore it. Then any of these things that result in decreased consciousness have the potential for causing respiratory effects.

I think one of the possible strategies that I'm trying to lead up to but I'm not sure I have a clear case for it is that patient self-report of reduced anxiety is really the therapeutic endpoint for this. We're not doing major procedures. We just want to get someone over the hurdle.

I have a trite observation. I'm now a high mileage kind of guy, and I've had about 12 things done in the last 20 years. I always say, "I just want local anesthesia and a little bit of oral sedation." And that works about 90 percent of the time, but if you can't get local anesthesia, my hand shoots up in the air. I say, "I want some fentanyl now or meperidine."

I even had a hernia repair done halfway awake, and the only thing that was disconcerting about that is when I was listening, I realized that my surgeon was on the left side and I knew my

Page 186

hernia was on the right side.

(Laughter.)

DR. DIONNE: That was anxiety provoking especially when some strange voice said, "Well, what should I do with this?" My surgeon said, "Oh, I wouldn't touch that if I were you."

Whoa, I didn't remember anything after that because my heart rate must have gone way up, and seemingly two hours later, I'm in the recovery room, and I'm looking down, and, "My legs are all there, but what about the other things that are down in that area?"

(Laughter.)

DR. DIONNE: All right. So if I'm accurate about this, sedation is really the observed manifestation of decreased consciousness. It doesn't necessarily translate into anxiety relief, although they are obviously correlated very tightly.

Then I got this at the last meeting, and I honestly can't remember who it was, but it was someone from the patient-reported outcomes office

Page 187

who said that we really shouldn't be using observer assessment because it's not a direct measure of how the patient feels, functions, or survives. So for a clinical trial's point of view, then maybe that makes a difference at least in my little shallow end of the pool.

I'm advocating then -- and we're looking at clinical trials but also change in clinical practices. Anxiolytic drugs are relatively selective so they make sense. The ability to titrate the dosage seems obvious, but when you have people that are using high doses of these drugs orally and they think they can titrate by waiting 10 or 15 minutes and then popping another pill down, that causes potential for problems.

The combinations obviously are prone to overdose, and then I think it's very problematic to have minimally trained dental assistants who are functioning as surrogates of convenience for the anesthesiologists.

Just to prove that this isn't all just hyperbole, I dragged up this old data. And this

Page 188

shows you 0.25 milligrams of triazolam compared to what turned out to be 18 milligrams of IV diazepam. We were titrating the patients to the usual endpoint of dropping eyelids, slurring of speech, and what we would call moderate sedation. And then we looked at the anxiety change from baseline, specifically asking about anxiety. Well, even with that good local anesthesia I'm telling you about, the patients definitely knew something bad was happening, and they had a big increase.

If you gave triazolam, you got about half as much anxiety report down to fairly low levels. Triazolam plus nitrous oxide, a little bit additive benefit, but the nitrous is so weak, when you put something really effective, it doesn't do that much. You can see nitrous alone, there's evidence at how weak it can be on its own.

Then diazepam, in a small sample of 10 patients per group, didn't actually achieve statistical significance, although it was obvious they were pretty well sedated. So it does suggest then a little dichotomy between -- oh, and the

Page 189

thing is these people look normal.
If you give someone 0.25 milligrams, unless it's time to go to bed, it doesn't affect them that much. Give 0.5, like I tried one time, you don't remember the post-op instructions. It can be used, and in some of these studies, we did do 0.5. If you give it sublingually, you get a faster onset, greater peak effect, and you even get patients reporting more pain relief.
So I think there's a difference between the anxiety relief and the appearance of sedation, which leads me then to some suggested endpoints and risk factors for outpatient sedation. It'd be nice to have some rigid criteria for percentage decrease or respiratory depression that we would say, based on clinical trials, based on these big footprints of deaths and whatever, we don't think this is what should be going on. And when we evaluate some new method, or if we ever get to the point where evidence-based dentistry is real, then we would say show us the evidence whether you have a safe procedure based on respiratory depression.

Page 190

I think it'd be a consideration to look at CNS depression as not good and the anxiety relief as the measure. And if you could have some way of measuring those two and show that there is some relationship between the respiratory depression, the CNS depression, and respiratory problems in morbidity and mortality, maybe that would be a reasonable endpoint.
Then many people talked about the status at discharge where people send patients out the door. I was reading one case report where the patient got a phenomenal amount of sedation, had a long procedure, and then died in the parking lot. The dentist tried to claim it had nothing to do with him, and it was just that person's time to happen and stuff like that, 10 minutes later. So there has to be some status for discharge as an outcome.
I think as a risk factor and I don't know how you measure this, I don't know how you legislate against it or whatever, but it seems to be logical that every place else in the universe, except for the dental community, thinks that having

Page 191

a separate anesthetist/anesthesiologist makes sense versus having a minimally trained dental assistant.
Even had one time we were interviewing someone for dental school, and we asked them if they had any research experience, and they said, "Yes, I do this anesthesia. I'm the dental assistant, and sometimes I experiment with which drugs I give and how fast I give them."
The captain of the ship was over there doing his procedure, not knowing that this little kid was squirting a little fast, a little slow, trying a couple drugs together and whatever. Imagine the maturity level of those people that are doing that kind of stuff.
Finally, one of the things that always strikes me is the range of response you get when you try to look across the population. It usually goes from the full measure. Whatever is zero and whatever is 100, you can show that when you give a fixed dose of a drug, you get a full dose of responsiveness.
It may be that part of the problem is

Page 192

because we always tend to treat for the worst case, we may be always picking the highest dose or a combination of drugs to try to achieve that outcome in everybody, where if we had some measure of individual response, then we could actually try to get to the point where we're just giving the safe amount of the drug to achieve the effective outcome. But again, it's hard to imagine other than for teaching purposes, but it'd be nice to arrive at that as something that we might try to capture in clinical trials.
Lastly, the safety of these multidrug regimens used for sedation, it appears to be particularly problematic in the pediatric population if over two-thirds of those deaths that I picked up were in pediatrics and then another five or six were in people that were extremely old or shouldn't have probably been in the outpatient setting anyway.
The young, healthy adults probably do okay because that's what they are, young, healthy adults who can absorb this stuff, but if we can get at

Page 193

some of this stuff, maybe that might be something to consider in our clinical trials' design. That's what I have to say about that.

(Applause.)

DR. WARD: We've got about 15 minutes before lunch, so maybe we can get the other speakers back up. And Randy, would you mind moderating the last session here?

Q&A and Panel Discussion

DR. CLARK: I'm going to take moderator's prerogative and ask the first question for Joe. Over the course of the development of the consortium, has the location of those sedation procedures changed? I know early on, it was highly ED specific. Is that still the case now that you're up to half a million?

DR. CRAVERO: No. I think the bulk of the procedures are done in sedation environments that are specific for pediatric sedation. So most of the institutions we collect data from have some location within their hospital where they perform a significant number of sedations.

Page 194

Now, the sedation team obviously needs to go to the MRI scanner, needs to go to CT scanner, et cetera, but most of the institutions do have a location where they perform sedations. We do not actually, within the data that I showed today, have many emergent sedations. There is some in there, but it's a relatively small amount that are done actually in the ED as an emergent sedation. There are sedation services that work within the ED environment, but they're doing elective sedations.

I'm not sure exactly how to answer your question. We do have a lot of data from emergency medicine specialists. It's second only to the amount of information we have from intensivists. And even from the intensive care perspective, when we collect information from intensivists, they are not largely doing procedures on patients in the intensive care unit that need an emergent procedure in the ICU. If they're performing the procedure in the ICU, it's because they have determined that a bed location in their ICU is going to be used for elective sedations.

Page 195

So it is a highly selective group, and I think that's one point we need to constantly bring up that it is very -- the groups that are participating are highly motivated and highly organized.

DR. CLARK: That's the lead-in to my real question. If I understood what you said correctly, you try to take context out of the reporting of events as much as possible, and if I understood that correctly, what do you think are the implications of removing context for the design of clinical trials?

DR. CRAVERO: Maybe the nuance of the language is not good from what I said. I think the context of what happens is very important, and again, we've talked about this for hours and hours and hours within our group.

I think what concerns me when you get to clinical trial reporting is that the idea of saying, well, this was a complication or an adverse event because it occurred in this particular environment, this was not an adverse event because

Page 196

it was another type of provider in a different environment gets hopelessly difficult.

I would advocate personally that when you talk about what goes on, it should be fairly objective reporting of what was done during these procedures performed with these medications to produce sedation and try to get away from having it be loaded with the idea of, well, because it occurred in this environment with this particular type of person, it should be considered this versus that.

That's more what I'm talking about. But I think when we talk about quality improvement and safety of patients, I think what you're pointing out can't be said enough, which is it's probably more important to consider who's giving the drug and what context they're giving it in than the drug itself. But again, I think that's different when you're talking about that versus clinical trials where we have a specific drug modality either given alone or in comparison to another.

DR. CLARK: Mark?

Page 197

DR. ROBACK: I really want to agree with what Joe said, and the contextual point that really matters --

(Laughter.)

DR. ROBACK: -- is the patients. When you look at Maala's studies and our studies when I was in Denver, emergency medicine for children, we are sedating ASA 1s and 2s, 99 percent they're receiving. They're healthy patients. They're receiving deep sedation but for really short procedures, whereas when you look at what these guys have published, 15 percent ASA 3s and 4s, they're getting MRIs that last 60 plus minutes. That's the difference, and you're going to see differences in your adverse events rates absolutely.

DR. LITMAN: I haven't heard much today about upper airway obstruction. Back in the 1990s, Denham and I did a series of studies that showed that when you sedate kids with very similar sedatives that they use in the dentist's office, what your real outcome is that's the most important

Page 198

is not really anything that has to do with ventilation per se, but it has to do with oxygenation.

Oxygenation really only goes down the tubes when you have upper airway obstruction. There's very little else that can cause it. You have to screen out for people, kids and adults, with upper airway obstruction, a propensity, like kids with big tonsils or kids with colds.

I oversee in my practice a very large amount of non-anesthesiology-driven sedation, and we're there to help out and to take over airways that become obstructed. Almost every time this happens -- in fact, I would just go so far as to say pretty much every time -- it's for one of two reasons: Either the kid had big tonsils or I should say some kind of pediatric sleep apnea that we didn't previously know about, but all you have to do is ask if the child snores at night -- it's usually a pretty good clue -- or if they had a cold, and they have some kind of upper airway inflammation.

Page 199

I don't really think it depends upon the ASA status per se, but it's the propensity to upper airway obstruction. So if you're going to talk about how to design a clinical trial, I really need to go back to like something that Denham and I talked about in the 1990s, which was how do you find a drug that has the best ratio of depressed consciousness to the ability to cause upper airway obstruction.

That's not an easy task, of course, because you have to figure out a way to measure upper airway obstruction, which isn't easy, and to my knowledge, the only drug that fits that favorable profile still to this day like in the '90s is ketamine. There's not much else.

DR. CLARK: Dr. Mason?

DR. MASON: I had a comment about dental because I think when we consider the dental sedation complications, it's a totally different kind of beast because even the data that you presented from Cote, that was before dentists were even using pulse oximetry.

Page 200

When you look, for example, in 2012, there was a closed claims database from a large dental malpractice insurer that released their adverse events, and of the sedation-related adverse events, half of them ended up in death, and only 8 percent of them had pulse oximetry.

Sp I'm not really sure that it's the drugs that we're giving the dentists, and I frankly think that the fact that the dentists haven't killed more people is because the drugs that we give them are relatively benign. Chloral's been around for a century, but I think it's more important to recognize that these patients aren't being monitored. There's no vigilance. There's drug overdoses. Patients are dying of just excessive lidocaine in dentists' offices or improper dosing rather than the drugs themselves.

DR. CLARK: We're going to have a discussion on the regulatory and practical differences affecting both medicine and dentistry a little bit later, but that's going to be an interesting part of that discussion.

Page 201

I think we had one over here first. No?
DR. KARAN: I also wanted to speak to the dentistry thing. As somebody who is training residents and has been asked by the school of dentistry where I am to train dentists, new recommendations to provide them some anesthesia and sedation training, it's very hard to relate anything that we're doing in the anesthesia world to anything that the dentists are doing.
I've asked them, "Well, why don't you tell me what you're doing in the community?" and there seems to be a disconnect. And maybe that will improve in the future, but certainly we can teach them to be afraid, as you said, and for proper monitoring.
As an anesthesiologist, I wonder if we're being mandated, appropriately so, maybe to teach them basic aspects of sedation, probably we're not using or modeling what we do for dentists to safe sedation for their training, for their requirements now that their national organizations are requiring.

Page 202

DR. DIONNE: The requirements have just recently been increased to 60 hours of didactic training and 20 cases for doing parenteral sedation presumably up to the level of deep sedation. For deep sedation and anesthesia, it's a more rigorous criteria. For the people who want to call themselves dental anesthesiologists, they have to do two full years of training.
The pediatric dentists have a -- I'm not exactly sure what their level of training is because they're usually assuming that because they're giving drugs orally, it's going to be okay. But you look at a lot of these things, one of the cases I found took place in a dental school clinic with a so-called dental anesthesiologist administering the drug, and quickly, when they realized they had a problem, transported the patient to the emergency room, and it was still a fatal outcome.
Even with those standards that they have in place now, that finite possibility or probability that something's going to happen when you're using

Page 203

certain drug combinations still manifests itself.
I like the idea of scaring them because I had a colleague years ago who told me he invited everybody in his hospital in the schools to come in and get anesthesia training as long as he was supervising them carefully. I said, "Geez, that sounds a little cavalier. I did three months of anesthesia, and the only thing I knew at the end is I was scared to ever do it again." And he said, "That's the idea, Ray." It might be something to that.
DR. CRAVERO: I'm actually going tomorrow to a meeting of AAPD to talk about data collection in a broader sense because there actually has been some legislation, particularly in California, that's going to require pediatric dentists to collect some information on what they're doing and report information on what they're doing. I think in conjunction with that, we may see some improvement overall in practice if we could actually understand what's going on.
The problem without a pediatric -- and

Page 204

correct me if I'm wrong, but a lot of problem is that we don't even know how many occur. There's no general reporting of how many kids are getting sedation and what is being used across the country, total black box as far as that's concerned. All we know is that every once in a while, there's a big problem.
If I could just put a plug in there, I think it's also important for us to recognize in terms of clinical trials that there are little or no clinical trials when it comes to office-based pediatric sedation for dentistry. Chris Heard and some people in Buffalo have done a couple along with the dental people there, but there's just not much there at all.
I think as a group of investigators and people that are interested in this, it's almost like something we should try to do more of is help pediatric dentists with clinical trials on the kind of meds that they use because right now, there's very little to guide them. We've done, I think, as a general population of researchers very little to

Page 205

help that.
If I could just push back on Ron just a little bit, I think what you're saying is absolutely accurate. We do have some studies that look at different populations, particularly obstructive sleep apnea since it's a problem that is rampant and growing.
I think we need to recognize that the problem of airway obstruction is going to be highly dependent on the population that we look at and that clinical trials need to report those comorbidities if we're going to make sense out of them.
What we will undoubtedly see, as has already been reported, there are certain drugs that are less likely to cause obstruction in a population of patients with obstructive sleep apnea, whereas, let's just put words to it, like propofol probably is less of a problem in 4-year-olds that don't have obstructive sleep apnea than it is in those that do. And part of clinical trials should allow us to understand what populations are most at risk and

Page 206

for that drug, for this particular trial, what are we talking about.
I would say, though, that just looking at things like snoring, we need to have much more precise definitions of what the population is and what we're talking about in order to make sense out of those clinical trials. Because as you know better than I, things like snoring sometimes can be extremely sensitive but not very specific for that problem.
I totally agree with you that the population and the comorbidities influence trial outcome, and we probably have not done as good a job as we should of defining those comorbidities.
DR. CLARK: Go ahead.
DR. LIGHTDALE: I'm just going to notice, I guess, that there is this inherent bias in asking the question did an adverse event occur. It's really in the eye of the beholder whether or not an adverse event occurred.
Mark, I guess my question with the TROOPS is, is there any thought to just collecting events,

Page 207

forget that first question, go straight into did any of these happen.
DR. ROBACK: I think that's a great point, and I was thinking about that as Joe was presenting as well. These are events of interest or events that we care about. It's just that we use adverse events for so long.
I don't know, Steve or Keira, do you have a thought on it?
DR. GREEN: I think just the idea of routine quality improvement, if you're tracking these lists of events, not all of them are clinically important. So you're going to burden your quality improvement process. The goal of TROOPS is to try and pull out what is clinically important and what's worth the time to track.
DR. LIGHTDALE: I'll just push back on that. If nothing occurred, it would take zero seconds to fill out the form, right? No events.
DR. CLARK: One last question.
DR. CHAPPELL: Phil Chappell from Pfizer. I'm sitting here listening to the conversation from

Page 208

the perspective of drug development. I've been struggling with this notion of the complex issues that have been kicked around, how do you define an event, and is it adverse or procedural related or drug related or some interaction between the two.
But I think that within industry, we would be forced to -- in an a priori way in a drug development program -- make some decisions or set some guidelines. An event of this nature would be recorded as an adverse event. It may not be related to the drug or the device under study, but I doubt we'd be able to have something of an agnostic description of the events that happened.
DR. CRAVERO: I would just say having dealt with the data monitoring boards in the past, clearly, you're going to have certain things that are unambiguous, and I think that's what I just tried to point out. From my perspective, the things that a data monitoring board would need to understand is that we have an unexpected admission when we have a kid who required ICU level care or was injured neurologically from a -- there's no

Page 209

question about that.
I think the question becomes when you have events that do not raise to the threshold of what an oversight board would need to know and adjudicate the continuing of a trial or not. That's a lot of what we get into, particularly in the pediatric realm, which people want to know how many times did you have to readjust the patient's airway to keep it opened, but I don't think that's something I would report to a data monitoring board. I would report if I had to call 911 to help me in my office.
But I get what you're saying. Certain things need to be not judgmental or not left to the provider, and we can be precise about saying it is a major problem. I think the question becomes in so many of the things that are reported in the pediatric realm are not clearly a big problem, and when you put language that makes it sound like it was a problem like "complication," it starts to get everybody uncomfortable. But I totally agree with you. There are major issues that need to

Page 210

unambiguously be called adverse events or complications.
DR. CHAPPELL: Right. Things like death or an unplanned admission are pretty clear and pretty distinct. The issues will arise, I think, in deciphering and deciding what to record for things that do not rise to that level. We may still within industry be compelled to record all of those events and place them in some category.
DR. CLARK: Dr. Ward? I'm sorry. Go ahead.
DR. ROBACK: I was just going to say that that's the big challenge because the obvious adverse events and outcomes are extremely rare, and then if we want to capture all these other events but we call them adverse events, is that going to decrease reporting because people don't want to -- this can become punitive, and I don't want them to think I did something wrong.
DR. CHAPPELL: Exactly. One last comment, and the person who spoke about MedDRA pointed this out. There's a distinction between an adverse event and what is now called an adverse drug

Page 211

reaction. For drug developers, it's the ADRS that end up -- a description of the risk-benefit of the product and so forth, or they carry the greater weight.
DR. CLARK: Dr. Ward?
DR. WARD: I think it's time for lunch.
DR. CLARK: Time for lunch.
(Whereupon, at 12:19 p.m., a lunch recess was taken.)

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

Page 212

(1:21 p.m.)
DR. WARD: In organizing the program, this morning was focused more on what's actually happening out there. So I had a little bit more of a quality improvement focus on how do we collect real world data to see what the adverse events are. Because if we're looking at something new in trying to design clinical trials, then we need to design them so that we're cognizant of what the real-world problems are that we're trying to make sure the new drug or technique might improve on.
This afternoon, I'm going to change the focus to really what we are interested in more, and that's if we've got a new drug or compound or procedure, how are we designing clinical trials that are best going to elucidate the true outcome safety issues that that compound might have.
We're going to start off with Leah from the FDA. I think what seemed to work the best this morning is we'll hold questions until the panel, and we'll get all three up on the panel. Frank's going to chair the panel, and then we'll have a

Page 213

break. Then we'll have a panel discussion that will encompass everything that we've talked about today.

Leah.

Presentation – Leah Crisafi

DR. CRISAFI: Thank you, and good afternoon. I get to do the after lunch talk which is always fun. I'm going to be providing regulatory perspective on evaluating safety and adverse events in procedural sedation clinical trials.

I've divided my talk into four sections. I'm going to start with the identification of drugs that are approved for procedural sedation. I'll spend some time talking specifically about midazolam because it is an example of a drug where serious safety issues were not identified until the drug was used in the clinical setting.

I'll then present the main challenges in evaluating safety in procedural sedation clinical trials. And I'll end with a few slides that include advice that we have given to companies developing drugs for procedural sedation.

Page 214

This is a list of drugs that are approved for procedural sedation and the year that they were approved. The first three are clearly indicated for procedural sedation, and whether etomidate, ketamine and methohexital have indications for procedural sedation may be somewhat debatable.

There are two reasons that I'm starting with this list. First, I wanted to point out that there does not appear to be much recent precedent in terms of evaluating and establishing safety of a drug for procedural sedation.

Second, I do want to briefly focus on midazolam, which might seem like ancient history, but I did not want to point out that not really a lot has happened in the realm of establishing safety of a procedural sedation drug since the time of midazolam's approval. And I do think that midazolam illustrates the importance of premarket characterization of a procedural sedation drug's safety profile.

It has been 30 years since midazolam was brought to market in the U.S., if I'm doing the

Page 215

math correctly, and so I'll refresh everyone's memories about that time by summarizing what was being released in the news at the time.

Midazolam clinical trials were conducted between 1980 and 1985 in settings where resuscitative equipment was available, and there were reportedly no deaths and no unexpected problems in the clinical trials.

In March 1986, Versed was first marketed in the U.S. and promoted as a drug for conscious sedation, and in 1987, the manufacturer issued two Dear Doctor letters, including a cautioning of the reports of deaths among patients who had taken Versed and the need for close monitoring of patients who received it. It was also reported that within 18 months after coming on the market, the FDA received 86 reports of serious adverse reactions, including 46 deaths.

The story goes on and includes a congressional hearing and criticism of both the company and the FDA, and a box warning for midazolam was added because of these adverse events

Page 216

that were occurring in the clinical setting.

This may be too small to read, and perhaps we don't need to read it, but this is the first paragraph of the box warning. It identifies midazolam as being associated with respiratory depression and respiratory arrest, and it does read, "In some cases where this was not recognized promptly and treated effectively, death or hypoxic encephalopathy has resulted."

I would like to make two points. First is the critical importance from a patient and clinician perspective of characterizing the safety of a drug that causes sedation, particularly because I think as we've already acknowledged today, sedation does often go hand in hand with cardio-respiratory depression. Second is how important it is for drug developers and regulators to strive to avoid repeating this situation, where the potential for a drug to reliably cause serious adverse events goes undiscovered in the clinical trial setting.

I am hopeful that our discussions today and

Page 217

tomorrow will be a step towards characterizing the safety of procedural sedation drugs such that we will not ever again what we as a community went through with midazolam.

Now, I'm going to move on to the challenges related to the evaluation of safety in procedural sedation clinical trials. So we have already hit on many of these challenges in the discussion this morning, and I look forward to continued discussion about the challenges.

The first challenge that I really wanted to talk about is the dynamic environment that is procedural sedation. During the course of a procedural sedation case, you often have changes in the level of stimulation, and those directly impact anesthetic requirement and cause changes in vital signs over the course of the procedure.

Positioning changes can also result in changes in vital signs, and those may be related to what I'll call effective blood volume such as when transitioning a patient in and out of lithotomy. Position changes may also result in blood pressure

Page 218

readings that don't reflect blood pressure at the vital organs such as during colonoscopy performed in the left lateral decubitus position with the blood pressure cuff placed on the upper or right arm.

Another challenging element in the procedural sedation environment, which also relates to effective blood volume, is the possibility of significant bleeding. Bleeding can cause changes in blood pressure and heart rate that are not related to any drug but rather a result of the dynamic setting of study.

Each of these changes reflect the dynamic environment, but is the clinical situation itself irrespective of any changes related to giving an investigational agent.

The next problem, if you will, is that your clinical trial investigators are likely to be experienced givers of sedation working in a very controlled environment and being very cautious because the drug in use has not been approved. They're going to be constantly anticipating and

Page 219

evaluating changes in patient status and taking steps to address those changes in order to prevent the occurrence of adverse events.

This is distinct, perhaps, from the clinical trial settings for most other types of drugs where you have a defined patient population and a defined intervention, that being administration of a study drug, after which the patient is followed and observed for adverse events.

So in the anesthesia setting, you have continuous evaluation and intervention that can mask or confound the identification of adverse effects of a drug that would be identifiable in the absence of that anesthesia provider who's doing their job to provide continuous evaluation and intervention.

The next challenge relates to the many data points that are collected over the course of a procedural sedation case. Most sedation drugs that we use have the potential to cause respiratory and cardiovascular changes, and one of the biggest conflicts in procedural sedation clinical trials,

Page 220

as I think we're already discussing today, is the distinction between characterizing a drug's cardiopulmonary changes and determining the incidence of cardiopulmonary adverse events.

This is a challenge because on one hand, we really do want to be able to inform clinicians about off-target pharmacodynamic effects of a sedation drug, and we haven't been historically considering any change in vital signs to be necessarily adverse.

We are constantly reconsidering the criteria for adverse until we do have well established and universally applied criteria for adverse. The most important thing may be the collection of complete data so that we have the ability to determine after the fact what to consider adverse.

Regarding the frequency of vital sign data collection, we've been trying to take a conservative approach, but it is not clear that a change at one point in time should be considered an adverse event. However, because we have been worried about missing transient but potentially

Page 221

important changes in vital signs caused by short-acting drugs, we have requested that sponsors document lowest values observed during the course of a procedural sedation case.

Ultimately, the minimum frequency of vital sign collection during sedation clinical trials is not established. Every 5 minutes is the American Society of Anesthesiologists' standard. Although it may not be a surprise to you that if we want sponsors to provide vital signs' nadirs, we have been less than satisfied with being provided data points for only every 5 minutes.

Perhaps this isn't a one-size-fits-all question. It could be argued that the pharmacokinetic profile of a drug be factored in determining the frequency of vital sign collection, or it could be argued that phase 1 and not phase 3 is the time for identifying pharmacodynamic effects of a drug as relate to basic cardiopulmonary function.

At this point, I will digress and share with you one sentence that I wrote during my first new

Page 222

drug application review in order to highlight our concern about the provision of vital sign data in a new drug application.

This is the sentence I wanted to share. "For example, this submission does not include a single blood pressure reading during a 100-minute adverse event of hypotension during which study drug was interrupted in a subject who ultimately died."

This is a clinical trial that was conducted in ICU patients, so the scenario is not exactly something we could imagine encountering in the procedural sedation setting. However, the concept is 100 percent applicable, that in order for us to be able to interpret what happens in the context of an event, be it considered by the adverse by the investigator or not, we need data.

Collection of this data needs to be incorporated into the study protocol and carried out by the investigator if we are to be in the position to evaluate and confirm adverse events.

Moving on to the challenge of concomitant

Page 223

medications, there are some situations where procedural sedation can be formed with only one drug being administered. Factors affecting whether one drug is sufficient would include the type of procedure and the pharmacodynamic effects of the one drug. However, it is much more common, as you all know, for several medications to be administered as components of procedural sedation, and those concomitant medications complicate the characterization of a drug safety profile.

There are two main issues stemming from concomitant medication use. Those are first, how do you ensure that the profile of the drug you are studying is reasonably well reflected in what you are capturing. In other words, are concomitant medications making a significant contribution to the safety profile because pre meds, rescue meds, and analgesics can significantly contribute to degree of sedation. They may be administered in significant amounts and produce sedation in which case, the safety profile may be more reflective of the con med than of the drug being studied.

Page 224

Arguably more important, particularly as we consider again the experience with midazolam and its synergy with opioids, is the need to understand the safety profile of the drug as it is going to be used clinically. If a drug produces sedation but provides no analgesia, it probably needs to be studied in the setting of invasive painful procedures requiring concomitant opioid administration so that we can understand the safety of the drugs in combination because their use in combination is inevitable if the sedation drug is to be used clinically at all.

Another challenge relates to the study of procedural sedation drugs in high-risk populations such as those with cardiopulmonary debilitation. We generally want drugs to be studied across the full spectrum of patients in whom they are likely to be used, and I would argue that it is important to include those who are debilitated to the extent that they may tolerate a general anesthetic.

However, this is obviously a very high-risk patient population, and challenges to study include

Page 225

the non-uniformity of the comorbidities in this patient population as well as likely difficulties in recruiting patients.

A final challenge that I think deserves mention but I hadn't been thinking would be our focus today, and probably we could spend more than an entire meeting talking about, is how to determine the safe setting for administration.

Our labeling for propofol, which I've chosen because it's probably the most used drug for procedural sedation today, states, "For anesthesia or monitored anesthesia care sedation, diprivan injectable emulsion should be administered only by persons trained in the administration of general anesthesia and not involved in the conduct of the surgical diagnostic procedure.

"Sedation patients should be continuously monitored, and facilities for the maintenance of a patent airway providing artificial ventilation, administering supplemental oxygen, and instituting cardiovascular resuscitation must be immediately available.

Page 226

"Patients should be continuously monitored for early signs of hypotension, apnea, airway obstruction, and/or oxygen desaturation."

My question that I will pose, but really may be for another day is, is there a method of evaluating a drug that would give us confidence that training in the administration of general anesthesia or resuscitative equipment are not required for safe administration?

At this point, I would like to restate what we find to be the major challenges with evaluating safety in procedural sedation. First, procedural sedation is a dynamic environment where changes are not necessarily attributable to the administration of a drug. Second is the continuous evaluation and intervention of the investigator who's doing their job by preventing adverse events. Third is the large number of data points that need to be taken into account in evaluating a drug safety profile.

Fourth is the issue of concomitant meds. If they are a major element of the anesthetic, then the safety profile may be more reflective of the

Page 227

con meds than the study drug, but if they're admitted from the anesthetic, then the safety profile established may not reflect the safety profile of the drug when it is used in the clinical setting.

Fifth is the challenge of studying high-risk populations where the establishment of safety of procedural sedation drugs is no less important than in ASA 1 and 2 patients. And last is the determination of the minimum requirements of the clinical setting where the drug is administered.

Now I'd like to move on to the final portion of my talk, which is just a brief presentation of a few fundamental pieces of advice that we routinely give companies relating to the evaluation of safety in clinical trials that I think this talk would be incomplete without. They relate to the definitions of what we expect to find in clinical trial protocols and the minimum number of subjects we require in a drug development program.

With regard to adverse event definitions, we expect sponsors to incorporate the definitions for

Page 228

adverse event and serious adverse events exactly as they are defined in the Code of Federal Regulations. You can see those definitions on this slide. The Code of Federal Regulations also includes definitions for life threatening, suspected, and unexpected, and ideally a protocol will also include these regulatory definitions.

There is a guidance that's listed here that we do find very helpful for identifying and explaining the definitions that we do often point sponsors to, and that's the guidance safety reporting requirements for INDs and bioavailability/ bioequivalent studies.

Regarding severity and causality determination for an adverse event, we expect protocols to include parameters for determining the severity of an adverse event as well as the relationship between an adverse event and the study drug.

With specific regard to severity, we usually point sponsors to the FDA guidance toxicity grading scale for healthy adults and adolescent volunteers

Page 229

enrolled in preventive vaccine clinical trials as a resource for severity definitions.

While obviously not developed for the procedural sedation population, we feel that it's a good starting place for sponsors who have not provided severity definitions that we think are reasonable.

For those of you who are not familiar with the vaccine guidance, I've provided this table as an example of definitions that have been used in the past and are found in the guidance. I want to point out that the identification of categories of mild, moderate, severe, and potentially life threatening as we have here is consistent with what we would expect a sponsor to define in their protocol.

Regarding causality determination, when a sponsor hasn't provided definitions that we think are reasonable, we usually point them to the World Health Organization Uppsala Monitoring Centre system as an example.

I've provided this table from the WHO UMC

Page 230

causality assessment system just to give you an idea of the causality definitions that have been used in the past. The WHO UMC example includes categories of certain, probable, likely, possible, unlikely, conditional unclassified, and unassessable unclassifiable. I would like to emphasize that we don't require companies to use the terms or the definitions provided here or in the vaccine guidance, but we do expect that companies provide reasonable terms and definitions that provide the basis for consistent classification within a trial of adverse events in terms of severity and causality, and ideally, the definitions are uniform across an entire safety database and drug development program.

Moving on to the numbers of subjects required for the demonstration of safety, we have told sponsors that as per the International Council for Harmonization E1A guideline, 1500 subjects need to be exposed to a drug that is a new molecular entity. We have also told sponsors that they must study a minimum of 300 subjects for each context of

Page 231

sedation with the majority of subjects exposed to the highest dose and longest duration for each sedation trial type. A final consideration with regard to the size of the safety database is the possible need for expansion if safety concerns arise during clinical trials.

With regard to non-new molecular entities, we have provided guidance that's very similar excepting the 1500-subject requirement. Companies have been advised of the need for at least 300 subjects per indication with the possible need for expansion of the safety database if issues arise during planned trials.

That is the last of the advice that I have to share. I'm going to just move on to a brief summary.

We first looked at the list of drugs that are approved for procedural sedation and established that there's not a lot of recent history that provides insight into the regulatory prospective on the safety evaluation of procedural sedation.

Page 232

I presented the example of midazolam where after five years of clinical trials, the risks of the drug seem not to have been well characterized. Then I presented the challenges in the evaluation of clinical trials, and I look forward to continued discussion about these challenges from the group today and tomorrow.

Finally, I identified some of the basic advice that we have provided sponsors relating to safety expectations in procedural sedation clinical trials, and that's it.

(Applause.)

DR. WARD: I think we'll save questions for the panel. We began with the segment on the regulatory perspective, and now a clinical trial design perspective.

Presentation – Daniel Sessler

DR. SESSLER: I've been asked to discuss clinical trials from the perspective of identifying complications. I'm going to address several different topics all bound together by the challenge of studying complications.

Page 233

Efficacy in a sense is easy to evaluate because efficacy is usually a continuous outcome. Furthermore, you have an efficacy outcome of some sort in every patient. Complications are very different because you don't expect them in most patients. They're inherently rare.

You can look at mediators of complications, so for example, hypoxemia as a mediator of respiratory arrest or vomiting as a mediator of aspiration pneumonia. So those events are a little more common. Some of them are continuous and therefore relatively easy to study or ordinal, and the reason that those are relatively easy to study is that you simply have more information than you do for a dichotomous event.

The trouble is that the events we care about, those rare but very serious complications, are always dichotomous. They're things like unexpected intubation, ICU admission, death. Those are rare and dichotomous, and it immediately gets you into trouble, and I'm going to illustrate how much trouble you get into.

Page 234

Studies are often powered for a 50 percent treatment effect or a 50 percent difference in the complication incidence in this case. But that's actually unrealistic because very few of our treatments actually have 50 percent types of effects. Twenty percent would be far more realistic.

If you take an event that occurs at, say, a 10 percent incidence, which is very, very high and fortunately, none of our serious events occur at anything remotely resembling 10 percent, but to design a study, a two-group parallel study, that identifies a 20 percent reduction in a complication that occurs in 10 percent of the cases, you need 5,000 patients. But most of our complications occur at, say, 1 percent, and then you need 50,000 patients, or 0.1 percent for serious events.

Things like death, ICU admission are probably less common than 0.1 percent, and suddenly, you're talking about half a million patients. So it is impossible to do randomized trials that identify these rare serious

Page 235

complications.

Perhaps as a consequence, our literature is full of fragile results. That's results that might be statistically significant, but don't actually give us very much information. I'll give you this as an example. These are two lightly disguised real studies, both of which were published in the New England Journal of Medicine, granted, two decades apart.

These were studies of a drug for prevention of postoperative myocardial infarction. One of these studies had 200 patients in it. There was 1 myocardial infarction in the treatment group, 9 in the placebo group, relative risk about a 90 percent reduction, and the p-value was 0.02.

The second trial had 4,000 patients. It had 200 events in the treatment group, 250 in the placebo group, relative risk of 0.8. That's 20 percent reduction in myocardial infarctions, and the p-value was exactly the same. It's 0.02.

Let me ask you, which of these do you trust? Well, the answer is obviously, you trust the second

Page 236

one, and there are two reasons for this. One is that a relative risk reduction of 90 percent is biologically implausible. Nothing we do reduces anything by 90 percent. It's not consistent with our experience or biology.

The other problem is that the result is statistically fragile, and what I mean by that is if you add two outcome events to the treatment groups in each of these studies, in the first study, you go from 1 to 3 versus 9. That result is no longer statistically significant.

If you do that in the second study, it does not change the p-value out to the third decimal. So that's a robust result. It's one that you trust. The first is not.

Let me put it another way. These are the results of theoretical studies, so we're reducing a 10 percent event to 5 percent in each of these cases. So it's a 50 percent treatment effect, already biologically probably implausible, but I'll give you that.

Each of these results is statistically

Page 237

significant, so you could publish any of these results, but let's look a little more closely at them. Look at the bottom one, for example. The 95 percent confidence intervals here range from about 0.25 to almost 1.

In other words, from a fourfold reduction in events, which is biologically implausible, to nearly 1, which is no effect at all. That study, even though it has 500 patients, this is not a small, inexpensive study. This is a 500-patient study, but it still is giving clinicians almost no useful guidance.

You do not know from that study over a factor of 4 what the true treatment effect is. You need 10 times as many patients. You need to get to 5,000 patients to shrink those confidence intervals to a level that actually gives clinicians some useful guidance about how to proceed that will allow clinicians, for example, to calculate number needed to treat.

Our literature is full of studies that are wrong or can't be replicated. The reason I say

Page 238

that is that if you look generally in the biomedical literature, about 90 percent of all papers report at least one statistically significant result.

If you look at very large NIH-funded studies, all of which are done for extremely compelling reasons, that is, good biological mechanisms, good animal data, usually some preliminary human data, so you look at those big studies, two-thirds of them are negative.

The difference between 90 percent positive and two-thirds negative is the error term. Those are all the publications out there that are wrong. The trouble, of course, is that we don't know which ones are wrong and which ones are right.

Almost everyone thinks that p equals to 0.5 means that there is a 95 chance of replicating the study. That is not at all what it means. P equals to 0.5 means that there is only a 5 percent chance that the observed distribution resulted from chance. That is not at all the same thing.

Let me show you what that implies for

Page 239

replication. Let's say we do a study of an intervention that is completely ineffective or we're studying a drug and it has exactly the same incidence of complications as your reference drug, same thing. You expect to confirm the null hypothesis. So you do the study, you expect to get a result near zero.

Let's say you then repeat the study, so exactly the same study, and you keep repeating it over and over again. On average, you will get zero because the intervention is ineffective, but you won't get zero every time. You will get a distribution of values around zero, and in fact, you'll get a typical, normal distribution like that.

What p less than 0.5 means is that the observed values from one study are in the outer 2.5 percent on each end of this normal distribution.

So let's say you do a study, and you get a p equal to 0.5. So the observed value is at the X there. That value then becomes your best estimate

Page 240

of what the truth is. You don't know what the truth is, but that's your current best estimate of the truth.

Let's do the same thought experiment again. So we keep repeating the study over and over again, exactly the same study over and over again. On average, you will get a value at the X, but of course, you won't exactly get X each time. You will again get a distribution of values around X, and in fact, it will be the same normal distribution just shifted so that the center is at the X.

Well, let's consider the implications for replication. So looking at the lower curve, half of these replication attempts will be to the right of the vertical line, to the right of the X value. Those values will more extreme than your original observation. The p-value will be smaller, and those will be considered successful replications. That's the shaded part there. But the other half will be to the left of the vertical line and the X. Those values will be less extreme and will have a

Page 241

larger p-value. All of those studies will fail to replicate the original observation.

So p equals to 0.5 does mean that you have a 95 percent chance of replicating the study. It means you have a 50 percent chance of replicating the study.

Well, 50 percent is a coin flip. That's not very good. A reasonable question then is, okay, how extreme a value do I have to observe in the first trial to actually have a 95 percent chance of replicating at p less than 0.5?

You get the answer to that by shifting this lower distribution to the right until 95 percent of it is more extreme than your original X value, the vertical line. And then you take the center of that distribution, and you go back up to your original, and you find out that you need a p-value of 0.0003.

Why on earth was p less than 0.5 identified as the threshold for statistically significant? It's really an accident of history based on a misunderstanding of what p-values mean. It never

Page 242

should have been the threshold. If the threshold for significance had been something like 0.001, we would have a lot fewer publications, and the publications we have would be a whole lot more reliable.

A strategy for dealing with rare events is to use a composite outcome. The reason people use composite outcomes is that it reduces your sample size by increasing the baseline incidence of events. Remember, sample size for a dichotomous outcome is determined by baseline incidence and treatment effect.

Treatment effect is part of the biology. You can't change that, but baseline incidence, you can change by broadening your definition of a complication or by stacking various complications.

The most common reason people use composites is to reduce sample size. That's actually not a very good reason for using the composites. There are compelling reasons to use composites, and that is when a particular disease or condition is manifested many ways.

Page 243

Diabetes is a perfect example. You're doing a study of glucose control in diabetics. You could design the study with a primary outcome of end-stage renal disease. That's legitimate, but don't you think patients are also interested in blindness and amputations? Wouldn't you want to include that? And myocardial infarctions for that matter, wouldn't you want to include that in your analysis? So that's a legitimate reason to use a composite.

The rule for a simple collapsed composite, that is, one or more, is that the components of the composite need to be comparable in terms of severity and incidence. For example, if you're evaluating, say, surgical infections, you could have a composite that includes deep sternal wound infection, sepsis, abdominal abscess, and urinary tract infection.

Whoops. Urinary tract infection is 50 times as common and is 50 times less serious. Effectively, all you're evaluating is urinary tract infection. So you're not allowed to do that. If

Page 244

you're going to use a simple composite, you have to find things that are comparable in terms of severity and incidence, or you need to use special statistical techniques, which are readily available that either account for incidence and severity.

Studies can be done with either superiority or noninferiority or rarely, equivalence. Most studies are done on a superiority basis. You want to see if a new drug, for example, for sedation is superior, that is, it's more effective or less toxic than an existing drug.

But let's say the new treatment is less expensive, or maybe it's less expensive and you have good reason to believe it's safer and you want to see if it's at least as effective. Then you might do a noninferiority analysis. Noninferiority is the same as saying it's not worse, and it's okay if it's better. Doesn't have to be better, but it's okay if it's better.

To do a noninferiority study, you have to set some clinically important delta because when you say not worse, you're not saying it's within a

Page 245

hundredth of a percent of identical. You're saying that it's within some clinically meaningful amount.

Let's say efficacy is defined on some sedation scale, and you say expect it to sedate to 3, but we'll accept 2.5 as clinically not different from 3. Then you would say anything that's about that good or better is okay.

Equivalence is rarely used. In fact, one of the few indications for it is evaluating generic drugs where -- correct me if I'm wrong -- I think the FDA wants the new drug to be identical, not better.

One way to enroll a large number of patients -- and if you're going to look at complications, you have to have a very large number of patients -- is this relatively new study method which I call alternating intervention. So far there's one published paper, one study completed, one that's in progress using this method.

It's not suitable for new drugs because it has to be done under a waived consent, but when you have two treatments -- let's say two standard ways

Page 246

of sedating people, but you don't know which is best or which causes fewer complications -- alternating intervention is a way of enrolling very large number of patients, many thousands of patients, relatively easily and relatively inexpensively.

It's appropriate for quality type studies where you have two interventions that are well accepted, they are both approved, and you have to convince your IRB that this is essentially a quality study, that you want to evaluate, to compare these two methods both of which are accepted.

If your IRB agrees, then what you can do is use one method for a period of time, say a couple of weeks, and then you switch to the other method for a couple of weeks, and then you switch back, and you keep doing that.

So it is not a randomized trial. Individual patients are not randomized to one treatment or another. In fact, the treatment periods are not even randomized; they just alternate. But you do

Page 247

this over a period of a year or two years or even longer. Over time, there is no reason to believe that patients preferentially get scheduled during one 2-week block versus another 2-week block. In practice, you end up with virtually identical groups, which is after all, the point of randomization.

What's nice about this is that you can enroll very, very large numbers of patients quite easily, and that's especially true if you're using electronic data acquisition, and all or most of your results can pull out of an electronic medical record.

I'd like to point out that even very well done studies that are technically done appropriately, they're blinded and randomized, can still give you results that are wrong. Attrition bias is not a big issue for sedation studies because people get their sedation and they go home. It's done.

But in, say, chronic pain studies where people need to participate for months on end,

Page 248

people getting less effective treatment drop out, and they drop out in a non-random way. It's called attrition bias, and that happens no matter how well you've done the study.

With novel techniques or novel drugs, the clinicians may not be as good. So they may be very experienced at using an older drug and much less experienced with a novel drug. In fact, if it's an unapproved drug, this may be the very first time they've actually used the drug. They're just not going to be as smooth with it. They're not going to know exactly how to titrate it. They won't anticipate problems the way they will with a drug that they've been using for years.

So you can end up with a result that the novel drug doesn't look as good as the conventional drug even if it's a better drug, and it's simply a matter of experience. The clinicians aren't as good at it.

Ancillary drugs may differ, and Leah mentioned that. If a new treatment is less effective, clinicians may make up for that by

Page 249

giving higher doses of ancillary drugs, and unless you're very careful, you won't necessarily trap that difference.

Then sedation levels may differ. So you may give doses of two different drugs, and they may even be set by protocol, but unless you have the right dose, you will get the wrong results.

Dose really matters, and let me illustrate that for you. So let's say you do a good quality, blinded, randomized trial. It's well powered. The 95 confidence intervals are small. Most people would look at this and say that's a pretty clear result. The experimental treatment is clearly better than the control. Anybody disagree?

(No audible response.)

DR. SESSLER: So these are the actual dose-response curves. Of course, you don't know the dose-response curve because for most of our drugs, we don't know the dose-response curve.

So those are your original results, and I've overlaid the actual dose-response curve. Notice that the dose-response curves in this case are

Page 250

identical. They are simply shifted a little bit. So the experimental drug dose-response curve has shifted a little bit to the left. It's a little bit more potent drug.

Well, let's say you used a higher dose of the control drug, which you might. After all, equivalent doses is not the same number of milligrams. It's some clinical impression about 4 milligrams of this is equal to 45 micrograms of something else. It's a clinical comparison. You're saying we think these are comparable doses.

But suppose you got it wrong. Suppose you had done the study a little differently with a just little bit higher dose on the control group. Suddenly, the experimental group looks substantially worse than the control group. But suppose you had given more of both drugs. Then you'd be up where the curve saturates, and they would look identical.

So my point is that dose matters, and we rarely include this in studies. Almost all of our studies have one dose of an experimental drug and

Page 251

one dose of a control drug, and you can get very different results depending on where you are in the dose-response curve. And this is a simple example because the dose-response curves are identical. They're just shifted. In fact, there's no reason why they should be identical. One could be flat compared to the other.

To summarize here, complications, at least the complications we're really worried about, are dichotomous and rare. Virtually no study is powered to detect serious complications. And as I explained right in the beginning, you essentially can't.

I'm not saying this to blame investigators. It's a function of the biology. When you're dealing with very rare dichotomous events, it is impossible to do studies that are large enough because we can't study 50,000 patients, much less half a million patients, in a prospective randomized trial.

Now, in phase 4 studies when you can use techniques like alternating intervention, then you

Page 252

can accumulate a large number of patients. It's important that drugs that get approved go into phase 4 studies. And midazolam is the perfect example of why we need to do that. You can't just approve a drug and say everything is fine because the rare events can't be detected in clinical trials. You will only see them afterwards.

Strategies that can help are composite outcomes, and remember that dose really does matter. Whenever possible, it is nice to include a dose-response curve in studies. Thank you much.

(Applause.)

DR. WARD: We'll save the questions for the panel.

Dr. Li from FDA will continue with some of the issues that Dan has raised about how we look at rare events in clinical trials.

Presentation – Bo Li

DR. LI: Good afternoon, everyone. My name is Bo Li. I'm a statistician from the Office of Biostatistics at the Center for Drug Evaluation and Research of FDA. I want to thank the organizers

Page 253

for this great opportunity to share and learn.

Actually, this therapeutic area of sedation drugs is new to our team, so this talk is pretty much a landscape talk. So I will share some general statistical comments on the quantitative assessment of drug safety. I changed the title later, a statistician's perspective working in FDA, and the standard disclaimer.

Evaluation of safety is a critical part of the drug review and approval process at CDER. I will give an overview of the safety evaluation of drugs and of the role statisticians play in that process. In particular, I will focus on some statistical considerations on these three items.

I think this is the wrong set of slides, but let me continue on that. So first, I will talk about the characterization of general adverse events reported in your NDA or BLA. I'll spend some time on meta-analysis of safety outcomes, then the challenges and features when designing a safety outcomes trial.

For the time consideration, I will skip the

Page 254

sentinel item, and I will talk about a case example for long-acting beta agonist, LABA, followed with some closing remarks.

Safety data is continuously evaluated at all stages of drug development, including the preclinical, early phase, and late phase trials. Before a new drug or biologic is tested in humans, preclinical work occurs to determine whether the product is reasonably safe for initial use in humans besides its efficacy.

The next step is clinical development. One goal is to get a safety profile for the drug in humans. Safety evaluation continues from phase 1 to phase 3 trials.

For marketing application of a new drug or biologics, FDA assesses whether the benefits of the drug outweigh its risks. Knowledge about a new product is always limited at the time of approval due to brief duration, limited patient population of clinical studies, or lack of sufficient information of some potential serious risk to be addressed appropriately in the product labeling. A

Page 255

new safety information often emerges after a product is used in a wider patient population after marketing.

In recognition of such limitations, FDA continued to monitor and characterize the safety of drugs through active and positive surveillance programs. With that being said, drug safety evaluation in FDA is continuing throughout the life cycle of a drug or a biologic.

In the last decade, several high profile concerns about drug safety led to the new regulation, including FDAAA, which stands for Food and Drug Administration Amendments Act of 2007. FDAAA granted FDA new authority to require postmarketing safety studies, and it changed the label to include new safety information.

Under FDAAA, postmarketing requirements, a PMR study can be required to assess the risk related to the use of a drug. That may be required at the time of approval or when new safety information arises. Such studies include randomized controlled trials, observational study,

Page 256

animal study, registry, et cetera. Before requiring a PMR study, FDA must find that a premarketing study is not sufficient. FDA must require at least a burdensome study.

FDAAA also authorizes FDA to require a risk evaluation and mitigation strategy, REMS, if it's determined either during the initial product review or at any point in the postmarketing period that specific safety measures are needed to ensure that the drug's benefits outweigh its risks.

FDAAA mandated the FDA create the Sentinel Initiative, an active surveillance system based on electronic health data. This surveillance system is called active because the FDA has the ability to initiate a query of the data.

A brief organizational chart of CDER. The Office of Biostatistics is under the Office of Translational Sciences and mainly collaborates with three offices in CDER: the Office of New Drugs, the Office of Surveillance and Epidemiology, and a relatively new Office of Generic Drugs.

Since the time that FDAAA became effective,

Page 257

FDA has substantially strengthened its safety program for drugs. Expanded groups dedicated to drug safety were established in CDER. In particular, within the Office of Biostatistics, the Division of Biometric VII, DB7, was created in 2009 to enhance the quantitative evaluation of drug safety. DB7 provides support to both the Office of New Drugs for premarketing safety assessment and the Office of Surveillance and Epidemiology for postmarketing safety assessment across the life cycle of drugs and therapeutic biologic products.

In DB7, we evaluate and help design safety studies, including clinical trials designed primarily to study safety outcomes. Such clinical trials could be either premarketing or postmarketing. We review observational studies submitted to meet postmarketing requirements.

When safety issues are raised by addressing the information, a retrospective look at multiple completed trials -- in other words, meta-analysis -- may be required, and a statistical analysis plan will be reviewed by us. We also

Page 258

review some prospectively planned meta-analysis to evaluate specific safety concerns.

In addition, DB7 has expertise in the design and statistical methods used in the sentinel studies and some other FDA initiated pharmacoepidemiological studies. When these safety studies or analysis are completed, we review the study report and look into the data and interpretation of the results. All these activities contribute to CDER's daily regulatory decisions. Besides the review work, DB7 conducts research of statistical methods in drug safety evaluation to support drug development and regulation.

Note that DB7 does not typically review the general adverse events of NDA or BLA. We get involved only when there is a focused or specific safety issue that requires the expertise and resources of DB7. But I will touch a little bit on some statistical issues arising in the general NDA/BLA adverse events reporting.

This table depicts the key differences of

Page 259

efficacy data and safety data collected to support a marketing application. Randomized clinical trials are the principal means of establishing the efficacy claims of drugs. However, these trials are limited in size and duration and exclude high-risk populations. Lack of statistical power and generalizability makes safety data included in an NDA or BLA mostly used for exploration and estimation purposes only.

The challenges also include the lack of prespecification and adequate ascertainment of adverse events. Safety endpoints are often not adequately collected, precisely measured, or adjudicated.

For evidence generation of efficacy, clinical trials are assessed individually. However, safety data are generally aggregated for multiple clinical trials. A reason to pool trials is that one may be able to provide a more precise and a more reliable estimates of safety parameters. Also, pooling data may allow conclusions to be drawn that would not be seen by looking at the

Page 260

individual trials.

The integrated summary of safety, SS, is a section of the NDA that provides comprehensive safety information collected throughout the drug's development program. The goal of the SS is to characterize the overall safety profile of the drug and to identify risks that should be included on the product labeling.

Safety parameters of interest typically include those specified in the FDA guidance, those that have priority, special interest, or concern for the compound or the drug class, and those identified during data review.

Some examples of safety parameters are exposures, concomitant medications, deaths, adverse experiences, lab measures, and vital signs. The summary of the estimates of correctly selected parameters should sufficiently describe the overall drug safety profile.

We can characterize adverse events by reporting crude proportions or incidence rates adjusted by exposure or time to event. That choice

Page 261

should depend on the trial design. For example, for time-to-event trial, proportions may be less meaningful.

Various methods can be used to make comparisons between groups. These methods include difference or ratio of proportions, difference or ratio of incidence rates, hazard ratios, survival curves, et cetera.

As we already discussed, pooling of safety data from multiple trials may give us more insights in the safety profile of a drug. From a statistical perspective, a critical question is how to pool data in scientifically sound ways. Rare adverse events pose additional challenges for data presentation in SS.

To explain the issues when pooling data from multiple trials, I made up this hypothetical example. Study 1 has two groups, treatment and a control. The randomization ratio is 3 to 1, 300 patients randomized to treatment group and 100 randomized to the control group.

We are interested in the association between

Page 262

the treatment and some adverse event, AE. Number of subjects with AE are 180 and 60 for the two groups, respectively. So it's easy to calculate the risk for each group. They are same, 60 percent. Thus the relative risk is a ratio of them, which is 1, means a neutral effect.

Now, assume that we have a second trial study 2, which investigated the same drugs and same outcome. This trial has a balanced design, 1 to 1. Each arm enrolled 200 patients, and each arm has 60 subjects with adverse events. So the risk of AE for both groups are equal again. It's 30 percent, resulting in a relative risk of 1.

We have two trials. Now, let's guess what will be the combined relative risk if we pool the data from the two trials. Intuitively, it should be 1 because for each individual trial, it's 1.

A typical pooling of the same SS is just crudely pooled across all trials. That means in the pooled table, number of subjects, number of adverse events are simply the sum of corresponding numbers of individual trials.

Page 263

Data are pooled together as if it came from a single study. Thus in this example, we got 500 subjects for treatment group, 300 for control group, 240 subjects with adverse events in treatment group, and 120 with adverse events in the control group.

Risk can be as easily calculated; again, 240 divided by 500, which is 48 percent, and for the control, that number is 40 percent. That ends up with a relative risk of 1.2. 1.2 means that the treatment is 20 percent more harmful than the control. However, this obviously contradicts with our intuition. It seems misleading.

This phenomena is called Simpson's Paradox. What caused that is deferring randomization ratios within a study and a different study populations across studies. Recall that study 1 includes a high-risk population with 60 percent subjects having the adverse event randomized in a 3 to 1 ratio. Study 2 include a low-risk population. The risk is 30 percent, and the randomization is 1 to 1.

Page 264

When we add the number of subjects from the two trials for those on treatment, 300 out of 500, 300 from study 1 out of the total 500, which means a 60 percent of the patients are high risk. For those on control, in total, we have 300, but you have only 100 from study 1 for the high-risk population. That means only a third are high-risk patients.

Crude pooling does not adjust for this disparity, thus resulting in a bias or distorted estimate of treatment effect. Crude pooling can give misleading results from any factor that impacts the adverse event proportion is disproportionally representing the overall drug and compared cohorts such as demographic factors like age, gender, race, or other factors like deferring time of study.

We can imagine a cardiovascular outcome. Two studies studied the cardiovascular outcome, and one is for younger population, and the other is for the older population. They have a different randomization ratio. If you mix them together

Page 265

crudely, that also will give you this target estimate of the treatment effect.

A lesson we learned is that crude pooling is not a proper way to combine data from multiple studies. We should almost always perform analysis stratified by trial. That means the overall estimate should be a weighted average of a common treatment effect or risk across trials.

Common traces or weighted method include the so-called inverse variance weighting and Mantel-Haenszel weighting. Both methods will lead to a point estimate of the treatment effect and its associated confidence interval.

We revisit this example. If we adopt the stratified or weighted analysis, no matter which weighting strategy we chose, we will get similar estimate of the overall relative risk with a point estimate of 1 as shown at the bottom of this slide.

We talked about the estimate of treatment effect when combining multiple trials. Let's now go back to the reporting of overall risk or proportions in the combined data. In this example,

Page 266

the overall proportion of the adverse event for each group as reported in this column highlighted in yellow, are 48 percent for treatment and 40 percent for control, respectively.

These proportions themselves seem misleading as one is higher than the other, but they should be comparable. If the proportions are comparable within each single study, an acceptable strategy should lead to comparable overall proportions between the treatment groups.

It's under debate which way is the best to report the overall proportions for multiple trials. One possible option is to estimate through weighting again. Two common weighting methods are Mantel-Haenszel or weighting by study size.

The overall risk estimated using the two weighting approaches were shown here in these two highlighted columns. Here is a type where it's not in my new slides, but that means adverse event, not death.

The overall risk estimated using the two weighting approaches, for one method is 43 percent

Page 267

for both groups, and the other method gives us 45 percent for both groups. They are both comparable.

For rare adverse events, very likely appropriate pooling is needed. Where events are rare, inverse variance procedure may not work well due variance estimate. We can consider other methods like Mantel-Haenszel or something called the Peto method.

Zero event trials are frequently seen in the setting of rare adverse events. In this case, the absolute effect measures like risk difference or rate difference may be better suited than the relative effect measures like risk ratio or rate ratio. Imagine you'll have zeros in the cells. You divide by zero, and that will give you indefinite number.

In many cases, while a formal comparison cannot be made, we can only report the estimate of the risk of adverse event and its corresponding confidence interval. When no events are observed, the rule of three allows one to calculate an upper bond on that risk. For example, in a sample of

Page 268

10,000 subjects with no adverse events, the upper 95 percent confidence interval for the risk would be set at 3 over 10,000.

I will attach meta-analysis of safety outcomes. If you search for the definition of meta-analysis, there are a lot of different languages. I personally like the definition given in this November 2013 FDA white paper. "Meta-analysis refers to the combining of evidence from relevant studies using appropriate statistical methods to allow inferences to be made to the population of interest."

Note here the keywords here are "appropriate statistical methods." Generally said, meta-analysis itself is a statistical approach used to combine results from different studies or trials to evaluate some specific hypothesis.

The stratified or rated approach is like when we have just the top [indiscernible]/ The inverse variance, Mantel-Haenszel methods, they are actually examples of meta-analytical methods.

Meta-analysis is often used when one single

Page 269

trial does not provide sufficient information, which is common for the rare safety outcomes. Meta-analysis can be used to estimate the treatment effect of risk for a therapeutic intervention and to quantify the uncertainty of the estimated risk.

By using all available data from multiple trials in your meta-analysis, randomization within each trial can be preserved. Statistical power is increased by increasing sample size. In other words, precision of the effect estimate can be improved.

Again, this is an older slide, so I have a lot of slides here for the meta-analysis, but I will talk about a general summary of the meta-analysis. I apologize for that.

We can do meta-analysis. More than one study has estimated a same effect. Choice of trials to be included in the analysis should be blinded to the results of that trial. One needs to evaluate appropriateness of study design and conduct of each trial, including randomization and blinding methods, patient population, outcome

Page 270

ascertainment, comparator, patient follow-up, and differential dropout. You remember garbage in, garbage out. The quality of the meta-analysis strongly depends on the quality of each individual trial included in that meta-analysis.

I'll skip this.

Assessed clinical trial information is unique to FDA. That means we can often get patient level data of the trials, which is ideal for a meta-analysis. The meta-analysis conducted outside FDA are mostly based on the summary results of the individual studies, which we call the trial level meta-analysis.

Patient level data allows us to apply common definitions of safety outcomes across trials to conduct subgroup analysis, to conduct time-to-event analysis to assess exposure and the follow-up between groups, to conduct various sensitivity analysis with all this detailed information.

To conduct a rigorous meta-analysis, selection of trials should be made blinded to the trial results. I emphasize that. We needed to

Page 271

assure adequate information was collected in an unbiased way, especially in a trial not designed for the outcome of interest. Many times data extraction from multiple trials are challenging due to inconsistent definition, collection, and the measurement of safety outcomes, and also due to the different structure or format of the trials.

I'll skip this, some methodology.

Based on what has been discussed so far, assessment of safety in drug development program has its unique methodological issues in the context of secondary use of efficacy of clinical trials in the context of rare safety events. Some rare events may not be even observed, and collaboration of information from multiple trials is often needed.

Meta-analysis is a statistical tool to synthesize the information from multiple trials. To do a high quality meta-analysis, you may need to team with necessary expertise, including statistical, clinical, or sometimes informatics. You may want to carefully develop a study protocol

Page 272

and a statistical analysis plan to conduct a rigorous meta-analysis.

Carefully designed and conducted meta-analysis can provide important input to FDA's regulatory decisions. In general, when FDA [indiscernible] for a prospective subject level meta-analysis.

Now I'll spend some time on the safety outcomes trial. Safety outcomes trial may be requested premarket or postmarket. The risk can be quantified only in a randomized clinical trial. Most clinical trials designed to evaluate safety are event driven, meaning the statistical information contained in that trial is determined by the number of events rather than the number of subjects.

Such trial is planned to continue following patients until a fixed number of events, let's say, D events, are observed. The trial objective is typically to rule out some amount of excess risk by comparing the upper bound of the 95 percent confidence interval against some prespecified risk

Page 273

margin, let's say delta.

The D events are observed by following subjects for a fixed period of time, and this provides the number of patient-years. We can imagine the highest annual baseline event rate is, the fewer patient-years will be needed to observe the fixed number events D. For rare safety events, we may need more patient-years, which means larger sample size, longer duration, so enriched population may be considered in such cases.

For example, in a dedicated cardiovascular outcomes trial, trials in which to observe patients at a higher cardiovascular risk would require fewer patient-years than trials conducted in low cardiovascular risk populations.

This figure shows the relationship of the risk margin and is the number of events needed when power is fixed at 90 percent, type 1 error fixed at 0.5, and assuming the true relative risk equals 1. In general, lower risk margin requires more events. As the risk margin increases, fewer events are needed.

Page 274

This is a table of results to show you specific values of the risk margin under the number of events. For example, if the goal is to rule out a relative risk of 1.3, you will need 611 events. However, if the risk margin is set higher at 2, only 88 events would be needed.

I have another table, but it's not shown here. I'll just describe it. So I have another table which shows you a different background event rates, like the rate is 1 percent or 2 percent, which can be considered as rare events.

For example, if you have risk margin of 1.3, you need 611 events. So if the background event rate is 1 percent, that means you need 61,100 patient-years. That is huge.

That means if the background event rate is low, the trial size to rule out excess risk can be quite large. That's setting a small risk margin for safety event that occurs infrequently would likely result in too large of a trial to be considered feasible.

In the end, the choice of a risk margin is

Page 275

highly impacted by the feasibility of conducting the trial and completing it in a timely fashion. Clinical considerations are necessary in how such trials will ultimately be powered as well as analyzed.

Another design feature that needs to be considered in a safety outcomes trial is a choice of a control arm. The choice of control can be placebo, background therapy, or standard of care, or even active control with a known safety profile.

We needed to consider knowledge of background risk of the control. For example, a control that has been under investigation for possible risk would not be appropriate. We need to consider tolerability of the control as it will be studied over an extended period of time. That's typical for a safety trial, also, ethics. For example, it may not be ethical to use a placebo control for a trial that is planned to continue for multiple years.

Safety outcomes trial included the rules for treatment discontinuation such as lack of efficacy

Page 276

after so many months of treatment or sustained increases in vitals. Additionally, this event-driven trial often has long duration that would result in fewer subjects being in treatment at study termination.

In order to assess the attributability of the event to treatment exposure, the trials should be designed to follow subjects while exposed to treatment as well as after the discontinued treatment. The statistical analysis plan should document how to address the attributability as well.

Study analysis includes all safety events that occurred while subject was exposed to treatment or off treatment. On treatment analysis, a subject is censored at the time of treatment discontinuation, plus typically, some predefined event ascertainment window.

Such analysis does not count events after the ascertainment window. These two analyses differ in how they count events in the defined time at risk. Overall, the assessment of the safety

Page 277

outcomes should include both analysis and prespecify which one would be the primary.

These are some examples of safety outcomes trial we currently see. We see a lot of cardiovascular outcomes trials associated with the use of antidiabetic drugs, which is due to this 2008 guidance.

I'll skip this. I'll skip sentinel.

The story of LABA, a little background. LABA is a drug class indicated for the treatment of asthma. They provide bronchodilation for 12 hours or longer. Some large trials conducted in 1990s suggested the LABAs are associated with adverse asthma outcomes such as asthma-related death.

This resulted in a box warning that warns of asthma-related deaths associated with LABAs and specify that these drugs should only be used for patients not adequately controlled on other asthma controller medications or whose severity clearly warrants initiation of treatment with two maintenance therapies. LABAs are currently used in combination with asthma controller medications like

Page 278

inhaled corticosteroids, ICS.

In response to the recommendations from the November 2007 pediatric advisory committee meeting, FDA initiated a meta-analysis to explore possible associations of four LABA products marketed in the U.S. with a composite endpoint of asthma-related hospitalization, asthma-related intubation, and asthma-related death. Another goal is to examine the risks in subgroups, particularly in pediatric patients.

In 2008, FDA requested sponsors of LABAs submit trial level and patient level data for asthma trials. These requests specified as the including criteria for trials the adjudication of asthma-related events, the format, and the variables of the data to be submitted to the FDA.

In the meta-analysis, the risk effect was estimated by Mantel-Haenszel risk difference, which is a stratification method stratified by trial. This statistical method makes use of trials with no events by using this risk-effect measure of risk difference instead of ratio.

Page 279

This forest plot shows the results of the meta-analysis for the individual drugs, four drugs here. Three out of the four drugs had a positive risk difference estimates. Remember, to the right of the Y-axis means it's bad for the drug. To the left, it means the drug is variable.

Three out of the four drugs had positive risk difference estimates for the asthma composite endpoint. Only one drug had statistically significant risk difference estimate, which is the second one from the top. The risk-difference estimate for one drug, the top drug, Advair, was negative and not statistically significant. Overall, you can see it's statistically significant for that risk difference estimate.

The meta-analysis results by the age subgroups is shown here. There was a general trend among the age groups with high-risk difference estimates among the younger age groups. Except the older equal to 65 age group, the risk difference estimates for all other age groups were positive and statistically significant.

Page 280

This trend among the age groups for the asthma composite endpoint was apparent when each drug is considered individually -- it's not shown here -- except in the case of Advair, the first drug. This kind of trend is pretty much driven by the asthma-related hospitalization. You can imagine that may be the most frequently observed adverse events in that composite.

Subsequently, the boxed warning of LABAs was revised accordingly to reflect this new information and current knowledge. This is for the pediatric patients. "Previous trials and FDA meta-analysis showed LABAs associated with asthma adverse events. It's not known whether there are similar risks when LABAs are added to ICS. Current available data are inadequate to determine that risk.

"It's determined that this question cannot be answered through re-analysis of existing data, analysis of spontaneous reports of adverse events, or epidemiological studies using existing databases. Therefore, controlled clinical trials are necessary."

Page 281

In April of 2011, the FDA issued a postmarketing requirement to all manufacturers of LABAs that are marketed for asthma in the United States to conduct controlled clinical trials to assess the safety of a regimen of LABAs plus ICS compared with ICS alone. The trials are multinational, randomized, double-blind and last six months. The primary endpoint is a composite of asthma-related death, intubation, or hospitalization. Events are to be adjudicated by an independent adjudication committee.

The agreed upon sample size of 11,700 patients in each trial will provide a 90 percent power to rule out a doubling of relative risk. That means [indiscernible] equals 2. The design and conduct of all the trials are similar so that the results of the four trials can be reviewed jointly in order to evaluate rare events such as asthma-related deaths.

To my knowledge, two trials have been completed so far, and the results have been published in the New England Journal of Medicine

Page 282

earlier this year. The study reports are under FDA's review now.

Some closing remarks, during the last decade, FDA has greatly increased its ability and capacity to address the quantitative safety evaluation of drugs through the successful implementation of new regulatory authorities of FDAAA and other key initiatives.

A safety system has been created to evaluate FDA-approved drugs across their entire life cycle. The quantitative assessment of drug safety focuses on premarket and postmarket safety studies for sound scientific basis.

Although great progress has been made, more work still needs to be done. For example, use of more refined data collection methods, encourage prospective planning and the design for safety assessment, and as always, the sponsors are encouraged to contact FDA early to discuss their research plans.

I missed the acknowledge and reference part again. So that's my presentation. Thank you.

Page 283

(Applause.)

DR. WARD: If we could get our three speakers back up for questions?

Q&A and Panel Discussion

DR. DEXTER: As the moderator, I'm going to start with the question, and then audience people can go from there. One of the things that struck me is that Dr. Li talked about in terms of a long-acting beta agonist, in terms of sample sizes of 11,700 for a trial, and it's definitely about 10-fold larger than what we're talking about in terms of the studies.

Dr. Crisafi, you mentioned in terms of one of the issues was the issue of the frequency of measuring vital signs.

Dr. Sessler, you talked about the whole issue that we really need to have sample sizes of 15,000, but practically speaking, it's going to be in the 1500-range initially.

I think that one of the things in terms of rather than a pessimistic view, but rather something we actually might be able to address, is

Page 284

the issue in terms of the frequency. So, for example, if you measure vital signs or record them every 5 minutes, that's a 300-second interval, it's quite easy to go from 300 seconds to 30 seconds.

One of the things I was hoping you-all might discuss, to think about the issue, is that if you're going to use the nadir blood pressure, the nadir saturation, the nadir respiratory rate as an endpoint, that's exquisitely sensitive to the sampling interval.

Is it feasible, as part of both the discussion now and thinking about it broadly as part of this meeting, to be able to come up at least with recommendations in terms of the sampling interval being something such as a 30-second range?

Do you want to start?

DR. SESSLER: I'll be glad to. I guess this is part of the general topic of what I could call curve descriptors because when you measure over time, whatever the interval is, you get a complicated curve. You can't possibly do comparisons over lots and lots of time. It's not

Page 285

meaningful.
You have to break down this complicated curve into something that you can describe either as a signal number or as some limited number of numbers. The simplest curve descriptor would simply be the average, but you could use the median. You could use the maximum. You could use the minimum. You could use the area under a threshold. You could do time-weighted average, time-weighted average under a threshold.
Any of these might be appropriate in various contexts. We've actually considered these in great detail because we've been very interested in blood pressure recently, and blood pressure is one of these things where you get lots and lots of measurements, particularly if you're tapping into electronic records.
So you can have hundreds to thousands of measurements per person times 500,000 people. You get lots and lots of numbers. How do you deal with them? We've actually looked at many different types of curve descriptors to find ones that are

Page 286

strongly associated with various outcomes of interest.
Now, the problem here is that we lack that association. We don't know, for example, the extent to which hypoxemia predicts things that we care about. Nonetheless, I think we can take frequent measurements, term them into a curve descriptor. Once you do that, it's normally not so sensitive to how many measurements you make.
Nadir is, and that's an exception because you get random values. Any single value is not perfect. Every value has an error associated with it as a technical error in many cases. And if you look for the lowest value of saturation, you will find some very low value that may have nothing to do with the patient, and furthermore, it may be maintained for, say, 3 seconds, which is not physiologically plausible or interesting.
With the exception of nadir, if you're using something like area under a threshold of 90, it actually doesn't make very much difference how often you measure. I'm a fan of measuring

Page 287

frequently, and I don't see any reason not to. There's no need to write these measurements down by hand. They are electronic data. They can stream onto a disk perfectly easily, and then you know exactly what happened.
You'll get the wrong values if you depend on people to write it down by hand. You will get wrong values in a non-random way because people looking at a complex signal like saturation that's going up and down all the time will pick a value they like and write it down. So I think you should just measure it electronically and then evaluate it with some objective electronic curve descriptor.
DR. DEXTER: Others? Any other comments?
DR. CRISAFI: I agree with everything that you said, and in terms of collecting nadir values, I agree that a drop to a certain level that sustained for only a few seconds really isn't clinically meaningful.
But we're very concerned about missing important things, and I think it will be great if we can figure out what those important things are

Page 288

that we need to capture and give us the nice buffer so we have good information, useful information, and not extraneous information.
DR. SESSLER: Right. That's the advantage of well-designed curve descriptors is that even if you don't recognize that a saturation of 63 maintained for 4 seconds is an artifact, it doesn't actually contribute very much to area under a curve because it gets bounced by everything else that happens.
DR. RIKER: Let me make a provocative statement. The ICU literature is growing with papers where surrogate physiologic outcomes, PaO2/FiO2 ratio, cardiac output, some other parameter of something bears no relationship to outcome or may even be opposite more meaningful outcomes such as functional evaluation, mortality, length of stay in the ICU, time on a ventilator, et cetera.
Let me challenge the concept that a single isolated vital sign ever means anything that's important to us as clinicians.

Page 289

(Laughter.)
DR. CRISAFI: We need to characterize a profile of a drug, and when you have a drug that's very short acting, if you ignore those dips as a drug is re-dosed, you may not have an accurate description of what the drug is capable of doing during a sedation case.
DR. WUNSCH: If you wanted to answer that before I go on a slightly different question, that's fine.
DR. SESSLER: I was just going to say the opposite is also true, that we've all seen patients who get what I call the dipsies. Their blood pressure goes down a little bit and recovers spontaneously, and this keeps happening, and then they crash. Okay? No particular dip was important. They recovered on their own, but it was, in fact, still meaningful. So there's no simple answer here.
DR. DEXTER: In addition, when we think of sedation, very often we're talking about the ASA 1 patient in the office-based setting as compared to

Page 290

the ICU patient.
Yes?
DR. WUNSCH: I just wanted to go back to the original question you posed using the example of blood pressure, whereas I think a lot of the answers were talking about pulse oximetry because that is a continuous monitor that can be downloaded as frequently as you want.
I just wanted to raise the point that when we're talking about sedation and monitoring, that to get anything more than every minute or two is going to start talk about having arterial lines in people, and when you're talking about thousands and thousands of patients, then you get into real risks that go with upping your monitoring to do that, whereas something like pulse oximetry and our other monitors are not invasive. I think it's important maybe that we make that distinction when starting to talk about how we monitor people.
DR. LI: I have a comment. Sometimes when we're looking at the vital signs to fully characterize the safety profile of a drug,

Page 291

especially for the safety outcomes we are not familiar with, we need to analyze such data.
For example, in the dedicated outcomes trial, you power the trial by the specific safety outcome, which is usually a hard endpoint like the deaths or cardiovascular events, something like that, for some hypothesis testing purpose.
DR. DEXTER: Dr. Cravero.
DR. CRAVERO: I just want to say this is like an awesome session. Really, all these talks were great. I'd love to steal all the slides, but I will ask some questions instead.
Dan, I was just wondering -- again, great talk -- one thing you didn't talk about was the difference between clinical and statistical significance. I think it was implied in a lot of what you said, but particularly where we have large studies, we can have large odds ratios with very little real clinical effect.
I personally see a lot of studies that I see published based on large odds ratio changes but with very little real clinical effect, and I was

Page 292

wondering if you would maybe just give your take on that.
For Leah, I had one question, too. I'm just going to throw out my questions. That is, you gave a definition that the FDA has for adverse events. It clearly doesn't jibe completely with what we've talked about here, and I'm wondering if you could give us an idea of what you think we need to do.
What is the FDA looking for from a group like this concerning that very important issue, which goes to how we study these things and do clinical trials?
Bo, I was just wondering, you made a real separation between efficacy and safety. I would suggest that in the field of sedation, those two things sometimes overlap. For instance, if a patient is moving wildly during a procedure, it could lead to safety issues, and therefore, I don't know that there's an easy, a bright line between efficacy and safety in this particular field. I'd be interested in your comments on that. Maybe Dan could start.

Page 293

DR. SESSLER: Okay. I guess the first one was for me, so I will start. This is an excellent point. There absolutely are clinically differences that cannot be detected statistically, and this happens in trials. That was the whole point of my talk, is that when you're dealing with rare events, you essentially cannot find them in any normal sized clinical trial.

On the other hand, when you go to epidemiological or history-based analyses, you have the opposite problem where it's very easy to find statistically significant associations that may not represent clinically meaningful effects.

It depends on what outcome you're looking at. If it's something like death, a lot of people would say almost any relative risk is important. But if you're looking at less important outcomes, that may not be true anymore.

Generally speaking, clinical trials suffer mostly from inadequate power and fragile results, but they're well done. They're internally consistent. Registry studies often find

Page 294

statistically significant results that are not clinically meaningful.

The real problem with registry trials is not that, though. It's confounding and bias, and they creep into unknown extents every time you do an observational study. I'm much, much more worried about confounding and bias than I am about statistical error in registry studies.

DR. CRISAFI: Regarding the definitions and identifications of things that are considered adverse, we have these prescribed definitions that are in the Code of Federal Regulations. We expect sponsors, companies to use the definitions that are codified.

I think we have the opinion that since we don't have thresholds that are universally agreed upon, or interventions that are universally agreed upon, as really clinically important, clinically significant, we feel like everything -- until we have some consensus about what really is and is not clinically meaningful from an adverse event perspective, till we have that consensus, we feel

Page 295

like everything probably should be considered adverse.

Sharon is raising her hand.

DR. HERTZ: I guess I actually just want to explore the question a little bit more because I'm looking back at the definition. Where do you think it is not --

DR. CRAVERO: I just think it's fairly general. We haven't been able to come to agreement in this forum -- and not that we've talked about it too long, but -- as to what represents a significant -- if you want to read it, the definition is just very general. That's my concern.

A group like this perhaps needs to help try to define how we should look at that definition because anyone of us -- we could take 20 people in this room. We read that definition, we may report different things because how we're interpreting what's written there.

What I'm wondering is what is the FDA looking for from a group like this to try to help

Page 296

further elucidate what they're talking about there.

DR. HERTZ: I don't know if someone is controlling the slides can put up Leah's slide -- I think it's 12 if I have the same version that came over.

There's no wiggle room on these. These are required by law in a clinical trial. From a clinical trial perspective, this would be a dumping of data, and that's okay. All the adverse events are expected to be reported. But I think the key here today and what you're saying is --

DR. CRAVERO: Can I just say, we haven't been able to agree on what's an adverse event.

DR. HERTZ: Okay. I'm thinking back to some of the anesthesia applications I've seen, and I'm understanding a little bit more now.

DR. DEXTER: If I may do this as sort of a moderator, you have a patient in which the plan is to give sedation during which they're going to be doing some sort of an upper endoscopy, some bronchoscopic procedure, in which it's totally to be expected that there will be hypoxemia

Page 297

transiently as part of the procedure. You just kind of then stop the procedure transiently.

Is that an adverse event when the saturation transiently goes below 90 percent? There is no practical way to differentiate between the drug and the procedure.

DR. HERTZ: Right. So I think the challenge here is to do a number of things. One is to figure out how to measure these events. For the purposes of regulation, everything is going to get reported. It needs to because -- that's a separate issue, but for the purposes of these studies and understanding the products, once you have decided how to measure them, then you need to decide -- hopefully, this group will have -- so there's the measurement, and then there's the relevance of it.

If you report every desaturation, you report every desaturation. It doesn't mean the drug's bad, especially if it's behaving in clinical context the way it's expected. In fact, it's determined that, for the most part, the background rate of hypoxia in the setting of bronchoscopy

Page 298

hasn't been exceeded. That's a discussion of the relevance of the things that are being recorded.

All those hypoxic events are not necessarily counted against the drug, and even better, if there's an active control, you compare them. And if there's no difference, the whole signal goes away. But it's still reported and discussed because what's important and interesting about that information is what if it's considerably less common with the new drug or more common with the new drug, then would be considered normal standard of care or as exhibited by the active comparator.

So the reality is even if the event of hypoxia is an expected event, in excess, it becomes an adverse event, and we don't know that until it's been recorded and considered in context. And that's where this group is important, is how does one do that. It's a two-step process. There's a measurement, and then there's an endpoint. And translating measurements into endpoints is much trickier in this context than in most because of the continuum between safety and efficacy.

Page 299

So that's the key. How do we gather enough information to understand what if the hypotension associated with the study drug is deeper than the hypotension with the standard of care?

So those are the kinds of challenges that we have in terms of getting so much data we don't know what to do with, wanting data that is feasible or not in terms of quantity, how to analyze it. These are the things.

So it's that intersect of coming out with measurements, outcome measurement instruments, and then possibly using that data, some other type of relevance instrument, where we want to quantitate it. That's where I think, like going back to this morning, perhaps that could help the difference between understanding a drug used in an outpatient suite for a procedure, in an inpatient suite for a procedure, in the OR, and in the ICU. It's all going to be context driven because that's how you guys will interpret these adverse events when you're using it, and that's how we need to know it behaves when we're looking at the overall balance

Page 300

of risk and benefit.

DR. DEXTER: Dr. Sessler?

DR. SESSLER: I couldn't agree more. All drugs, all interventions, everything we do has potential complications. The outcome of a randomized trial is not that there are complications. It's the difference in complications between the two groups and therefore, it's perfectly okay that a procedure like bronchoscopy causes hypoxemia, but if one drug ends causing a lot more hypoxemia than the other, then that's interesting.

Along those lines, I think it's helpful to predefine clinically meaningful differences, and that's something that investigators are beginning to do, but it's not actually been routine in the past. People would just look for a difference and hope they find some statistically significant difference. And if they do, they write a paper about it saying, okay, there's a difference.

If you predefine a difference and then you end up with a small difference -- and this happens

Page 301

in pain studies all the time. You end up with a difference of 1 point, 2 points out of a 11-point Likert scale. Is it clinically meaningful even if it's statistically significant? Probably not.

DR. DEXTER: Yes?

DR. CHAPPELL: I have a follow-on to this comment. We routinely, or in many trials at least, in addition to collecting all the adverse event and vital sign data we're required to collect, we'll predefine adverse events of special interest, and they will often have criteria for what is held to be or considered a clinically meaningful effect. Sometimes it might even have requirements that patients be terminated from the study or other steps taken if the situation arises.

That might be one way to address the need. On the one hand, it would be comprehensive to collect all this data. On the other hand, to be able to target effects that are likely to be clinically relevant and meaningful.

DR. DEXTER: Why don't we finish here, I think. Yes, please.

Page 302

DR. LI: My comment to the question for me, I think Sharon and Phillip already addressed a lot. I think in some contexts, you have a clear cut of safety and efficacy, but in some contexts, if you don't understand the drug very deeply, then you don't have a clear cut of that.

For example, you have common adverse events that may be expected for some compounds for some populations, and then sometimes you have the adverse events of special interest that was defined a priori, which you may know a lot or may not know a lot. And sometimes you have unexpected serious adverse events.

For example, that Avandia story, that's unexpected. I think the cardiovascular harm, which was shown in that meta-analysis, was unexpected. That's why FDA has this 2008 guidance for the anti-diabetic drugs to evaluate their cardiovascular safety.

Now, we are seeing some cardiovascular outcome prior for diabetes drugs powered for superiority. So that means the sponsor may want to

Page 303

test if we have cardiovascular benefit.

So I agree to some extent. It's a continuum. Sometimes there is a clear cut. Sometimes there is not.

DR. DEXTER: Do we stop --

DR. WARD: You can do some more questions.

DR. DEXTER: Yes, you have a question?

DR. HERTZ: I just want to ask a question about the minimum clinically important difference. I am more familiar with analgesic studies than anesthesia studies because, frankly, we get more.

We often see a group mean difference in treatment effect that's rather tiny. You said you were questioning the relevance of a 1.1 difference on an 11-point scale, and sometimes we see a group mean of difference of well less than 1, 0.5, which is pretty big for most of our studies for a variety of reasons. But I don't think a patient would ever say, "My pain is down a half a point. I'm feeling a lot better."

So I think what's really important when we think about clinically important differences is to

Page 304

separate what we mean on an individual basis and what we mean on a group treatment difference.

For instance, with blood pressure, when we see blood pressure studies and we see a difference of 2 or 3 millimeters of mercury, that's considered pretty big on a population scale. But again, that's within the range of noise for having the person rush in a little bit late or even if it's just the normal fluctuation. We would never make a therapeutic decision based on 2 to 3 millimeters of mercury on an individual.

So as you think about how to put these measurements into context, it depends how you choose to look at the data. If you look at average changes and you think that's relevant, then that's what's different, what's meaningful from a group perspective. If you're looking at responder definitions in individual amounts that count as useful, and then you're going to count the people who have a useful or whatever change, that's another way to look at it.

I think it's just important for us to

Page 305

remember that there's two ways of looking at minimum important clinical difference, individual versus group.

DR. SESSLER: That's certainly true. Some of these may be the wrong types of studies. Maybe they should have been noninferiority studies to start with because it sounds like you're getting that kind of result, and a new drug that's noninferior to another one may still be preferable under some circumstances.

DR. DWORKIN: I disagree with you. If you've got a drug where you've replicated statistically significant superiority to placebo and the delta is 5 out of a 100, 5 millimeters on a 10-centimeter VAS, and this drug is very safe, very well tolerated, has a novel mechanism of action, and is relatively inexpensive, I would argue that's a contribution to public health.

There's no threshold for what is clinically meaningful at the group difference level absent a consideration of all of these other factors like safety and tolerability and cost, novelty, et

Page 306

cetera, et cetera, et cetera. A 0.5 delta, or 5 millimeter delta, obviously is not clinically meaningful if the drug is less well tolerated and riskier than what's on the market.

So the group difference that Sharon was describing, we don't have any thresholds for pain, which is all I know about, because the context of that delta is so important, primarily safe in tolerability, with all these other factors.

DR. SESSLER: Right. I was making an argument there. But sure, a new drug that cost 10 times as much with an unknown safety constellation has to be a lot better than just equivalent. On the other hand, in -- I do lots of thermal regulation studies.

If you have a new warming device that's less expensive, you have no reason to believe that it's dangerous, I would say all it has to be is as good as our current warming systems. What approach you use is very much dependent on the circumstances.

DR. DEXTER: Yes?

DR. WARD: Just a follow-on that you gave

Page 307

the illustration that we need a dose-response curve when we're looking at the effect. Unless we've got a single dose, you can't really tell. But you also need to have a dose-response curve for the adverse events, because if you don't have the dose-response curve for the adverse events, you may find two drugs that look better that fall into with what you just said. You really need both dose-response curves.

DR. SESSLER: Oh, absolutely. The dose-response curve is going to be different for every effect of the drug, and you have to look differently at different effectiveness measures and separately at different adverse events. We never do this. So the dose-response curve, it's very easy for me to draw them on a slide, but in fact, we don't know what they are.

DR. DEXTER: Let's take the last question. TJ?

DR. GAN: I think one of the problems with what I'm trying to raise is that I think the instrument of measurement that we have, it's very

Page 308

rudimentary. We talk about the VRS score and 1.0 to 10. The fact is that patients interpret very differently. So a score from 9 to 6 or 7 is very different from a score of 4 to 1, although the extent of the difference is the same, but it's very, very different.

So in a way, we haven't really taken into account what the individual patients think about the drugs. We just look at because of our constraints and limitations of the score that we have. Another one is the nausea score from zero to 10. Again, we know that that is very different, and patients interpret it very differently.

I think the whole thing may be a little bit flawed because we just don't have good instruments.

DR. DEXTER: Let's end it there, and say when it comes to efficacy measures, that different scales have advantages and disadvantages in terms of their interpretation perceptually.

DR. WARD: Let's take a half hour break, and we'll come back with a panel. We'll wrap up for the day.

Page 309

(Whereupon, at 3:23 p.m., a recess was taken.)

Panel Discussion

DR. WARD: I thought we'd spend the last hour today -- from my perspective, I think this has been an incredibly productive and interesting meeting. I hope it all has been for you. Tomorrow we will focus more on some of the individual adverse events.

One of the key things that I heard today that I really liked as an organizing event -- I think it was Susan that said it -- is we really have two instruments. We have a measurement instrument that's going to be collecting the data, and we have a relevance instrument that may be context patient and provider specific that then filters the measurement data to decide what the relevance of it is for particular adverse events. I like that concept of thinking about how we measure what the relevance and incidence of adverse events are.

This is the time for anything we've talked

Page 310

about today, both comments from the audience, questions and comments from our panelists. I want to start out as the moderator with a couple of questions for the panel, but for everybody, too.

The first one is, if I'm doing a clinical trial for procedural sedation, and particularly looking at adverse events in it, should a blinded control group always be included in a clinical trial of a new procedural sedation, or should I -- many of the clinical trials I see for procedural sedation drugs are, let's just try the drug in a set of procedures. We'll measure its efficacy, and we'll measure its adverse events, as opposed to a randomized controlled trial in which we've got a control group. Clearly, it can't be a placebo. It's got to be an active control. If so, if we should be recommending that, what's the active control?

So I'll turn that over to anybody in the panel who wants to --

DR. SESSLER: Well, to be a trial, it has to have a control group. What the control group

Page 311

should be, it depends. If you have no available treatment for a condition, it's acceptable to use a placebo; otherwise, most people would say you should use your best available treatment as the control group and then compare your novel entity to that.

DR. WARD: What's the best available treatment?

DR. SESSLER: What is best?

DR. WARD: What's the best available treatment in a clinical trial for procedural sedation?

DR. SESSLER: I guess the investigators and clinicians can decide that, and it's going to be a highly context specific answer.

DR. CRAVERO: I would sort of agree. I wrote an editorial about trying to raise the bar for clinical trials in pediatric sedation a while ago. I don't think there was anything profound about it, but I think it goes a little bit to this, which is, number one, in pediatric sedation trials, it is often just as you described.

Page 312

They're called trials, but really, it's an observational report of the last 100 whatever I did with whatever drug I used. And they're not terribly helpful, yet that's what gets reported, as I think I tried to outline a little bit today.

I do think what was brought up earlier today which is awesome, is that there is a real dose issue as well, and the comparison is always made with whatever the investigator has chosen as the comparative drug and dose. Yet oftentimes, I would look at the trial and say, that's really an inadequate dose.

There's a million different examples we could all use, but we know, for instance, the efficacy of a drug like dexmedetomidine at a certain dose in children is pretty low, yet there are reports of it used at much higher doses.

I would throw it out there. I don't know what the right answer to that particular question is, but if you're going to use a low dose of that drug and compare it to something else, you're going to get one result, whereas if you used what has

Page 313

been reported as a much effective dose of the comparator drug, you're going to get a different result, and Dan reviewed that very nicely.

I think there should be a comparator group, and there should be a comparator group that has some data behind it that reveals that that is an effective way of using that comparator drug.

DR. WARD: Hannah?

DR. WUNSCH: You mentioned the word "blinded" in your question, and I think that raises all kinds of issues in terms of kind of handcuffing providers who may be, for instance, comparing sedatives with very different qualities, and where it really is just not practical, or really safe in a sense, to be asking providers to be titrating things. And this gets back to the dose issue where often you're not really sure whether doses are equivalent or not.

So although I completely agree that to say that it's a trial is to have two arms, I'm not sure blinded needs to be in there, and maybe shouldn't for some of the safety reasons.

Page 314

DR. WARD: Any comments from the panel?

DR. SESSLER: It's often not possible or practical to blind a study, but let me just clarify the wording. If it's a trial, you have to be comparing two different things. Normally, that's randomized, blinded if practical. It doesn't have to be randomized. So an alternating intervention study is a type of trial even though it's not randomized.

A case series is not. That's an observational study. A retrospective analysis is an analysis. It's a study, but it's not a trial.

DR. WARD: Rich, and then Jerry.

DR. RIKER: I guess I would say I don't know that we should recommend a single drug or a single approach to procedural sedation because the specific qualities that we're trying to attain are so different from procedure to procedure. Movement okay, yes or no. Recall, yes or no.

I like the idea of looking at the available evidence and looking for a proven comparator, one that many people would agree is an accepted

Page 315

standard. But I don't think we should go into the weeds of making a drug recommendation.

DR. WARD: Jerry?

DR. LERMAN: Two points. The first is, unfortunately, many of the drugs we have for use in anesthesia, except perhaps the inhalational agents, have never had a dose-response study actually performed on the drugs for any outcome, efficacy or otherwise. So before you even get into the sedation realm, the expected effect of the drug has never been studied.

So you do have -- and I see it on editorial boards that I sit. We get all kinds of people submitting papers, and they picked arbitrary doses because everybody else uses them. However, that's all flawed. And the primary problem is we've not got those dose-response studies, and we should be doing them as part of the FDA approval of the drug.

The second thing is I beg to differ with those on the panel who say you can't blind any of these studies because the notion that you don't know what drug you're giving isn't actually

Page 316

necessary to blind a study. The observer, if they're recording an effect, need not know what the hypothesis of the study is or what their actual -- the key elements and the outcome are. They can record all kinds of extraneous data.

If they don't know what you are actually seeking, then actually the study is blinded. Now, if the operator is also making those judgments based on the drug that person is giving, then that makes it more complicated, so you have to have a distinct observer. That person unaware of the hypothesis or specific issues that you're looking at makes it a blinded study.

DR. WARD: Yes, I think those are good points, Jerry. A couple of studies that I've done in the past, for the original study on dexmedetomidine, we actually did a dose response for dexmedetomidine, both on effect and side effects, so there are some things in the literature with that.

DR. LERMAN: But not in children. That's the whole -- that's what we're talking about.

Page 317

DR. WARD: In children, okay.
DR. LERMAN: With inhalation agents, we did determine MAC values in all the agents before -- well, halothane was on the market for 10 years before George Gregory found the math for it, the measure. So we do have that disadvantage.
DR. WARD: It is possible to do the blinded study. I've done a study in which I was giving the drug behind a shield, and nobody else knew what drug was being given because they couldn't see it. You have to do a few things, but I think it is possible to both do dose response and blinded.
DR. MASON: I think it also depends on the drug. If you're comparing ketamine to anything, it's not going to be valuable to blind because you're easily going to tell, or dex, it's going to cause drop in heart rate.
So I think the sedative drugs in general seem to have different enough properties that it would be really not possible for somebody not to guess what they're giving.
DR. LERMAN: If they actually don't know the

Page 318

drug, they can guess as much as they want, they're blind.
DR. WARD: Yes. I would just say that there's a big difference between guessing and knowing, and you'd be amazed at the number of times the guesses are wrong even though you think they should be able to tell what the drugs are.
Comments from the panel?
DR. SESSLER: I agree. People are not nearly so good at guessing the drugs as they think they are. Observer blinded is a good way to go if you can't blind everyone, but the more people you can blind, the better.
We typically keep our statisticians blinded, also. So they do their analysis on a group A/group B basis. So you might think statisticians are completely objective, but they're making decisions all the time. Is a value an artifact, or is it real? How are we going to do a particular analysis? We just keep them blinded.
DR. WARD: I don't know what the study was. Oh, it was on using sham orthopedic surgery versus

Page 319

I think it was for knees. And that was so blinded that the authors wrote two versions of the paper before they knew the results. They wrote a version of the paper if it was a positive result, and they wrote a version of the paper as if it was a negative result before they unblinded the results. Not only were the statisticians blinded, but writing the paper was blinded.
DR. SESSLER: We often write the paper before the results are available.
(Laughter.)
DR. WARD: They actually wrote two papers. They actually wrote and agreed upon two papers, one with a positive result and one with a negative result, and agreed that when they unblinded and got the result, they couldn't then change the paper. That was the paper that they were going to have to submit.
DR. SESSLER: That was a classic article.
DR. CRAVERO: I obviously totally agree with the comments. I would only say that lacking good PK/PD data on all of these sedatives in the next

Page 320

10 years, when we see clinical trials, it would just be nice if people would recognize that what they did was kind of arbitrary and try to use the available evidence, such as it is, to choose a reasonable comparator. I don't think that's always done. That's my only point.
DR. SESSLER: I would love to see the FDA require good dose-response curves before drugs go into clinical trials, that it go from a very small phase 1 dose escalation study into a formal dose-response curve, and then go on to phase 2 and phase 3 studies.
DR. HERTZ: Me, too. I would like to require that, too.
DR. WARD: Susan, you want to make some comments? Go ahead.
DR. HERTZ: I would love to have the ability to require that.
DR. WARD: Mark, and then Albert.
DR. WEISS: At the end of Dan's talk, which was really wonderful -- Dan, and I talked to you about some of that, too. I want to hear the

Page 321

group's thought about this, too.
Dan talked about a new drug that might really be more efficacious but it costs X times more, so what's the value there. And I'm wondering if we need to consider economic considerations as part of maybe efficacy because when you're in private practice, they might say that a certain drug is better, but the hospital will tell you, you're not going to get that drug because it costs this much more.
I'm wondering if maybe this is part of the conversation that we have as well, or maybe the conversation will be eventually made for us.
DR. WARD: Albert and then --
DR. DAHAN: I think the idea of doing PK/PD studies, and I wonder whether you should do two dose-response curves because we often titrate to effect. So if you keep measuring your plasma concentration, you measure your effect. You do have in every patient that you test a dose response. When doing that, you don't need that many subjects at all to get a good result.

Page 322

DR. WARD: Comments?
DR. DEXTER: I want to just comment on the issue of the economics. When it comes to the sedative agents, one of the -- when I say challenges, I don't mean a negative or positive; it's just one of the issues -- is that very often, the economics is more dependent on the context of use than it is on anything of the property of the drug.
For example, if you have a particular gastroenterologist, pediatric gastroenterologist on a particular day, and there are three hours of cases in the operating room, it makes no difference if one drug is faster than other. It's a fixed cost.
In contrast, if you have exactly the same drug with exactly the same profile, and now you put it into a room where the pediatric gastroenterologist has 10 and a half hours of cases. In that circumstance, it's probably purely a variable cost.
One of the issues in terms of considering it

Page 323

is separately the issues in terms of timed reductions, which can oftentimes be done at the same time as an efficacy study, although you have to consider as part of the trial whether it's a realistic measure of time differences.
That's different from the economics of the drug. So therein lies one of both the features of sedative agents as well as different anesthetic agents, is focusing really on time as an endpoint.
DR. SESSLER: Cost-effectiveness studies age quickly because costs change quickly over time, and they may also be very different from one hospital to another. Different hospitals with different bargaining powers simply pay different amount for drugs.
DR. DEXTER: That's why you measure the time difference as an endpoint, which don't really age. They can somewhat differ depending upon the workflow, but generally are very stable, homogenous among centers. The economics of it differ dramatically, not only among hospitals and over time but just between two different operating rooms

Page 324

in the same hospital and the same day.
DR. CRAVERO: I do think it's an interesting question with respect to safety, though, concerning the specific goals of this particular conference, or this particular gathering, which is if we have drug A that is clearly more costly from the perspective of the time that it takes to recover or other aspects of the drug, the actual acquisition costs and everything else as Frank just said, that would go into cost analysis versus another drug. Yet by the definition of some of the things we talked about this morning, it has more requirements for airway repositioning or more desaturation episodes, yet no meaningful outcome differences.
Do you say, well, how do I consider that? I'm not sure what the answer is, but we haven't discussed that. If the drug was three times more expensive but had fewer desaturation and airway repositioning requirements, how do you put that into context? I don't know. And it probably does age fairly quickly.
DR. DEXTER: I was going to say also, airway

Page 325

repositioning depends in part on the level of the provider. Are you referring to a pediatric anesthesiologist? Are you referring to an assistant in a clinic? So that has an enormous effect in terms of thinking about these adverse events.

DR. WARD: Back row.

DR. TOBIN: Having been a previous FDA advisory committee member, part of our introduction to the axioms of the FDA were to ensure the public safety. It had nothing to discuss about pharmacoeconomics or cost analysis.

So I think in a regulatory environment, remember, the FDA doesn't set the price. The FDA doesn't recommend what the competitive advantage price is. The FDA is to ensure the public safety. The sponsor decides on the price and what kind of premium discount you're going to get if you're buyer A or buyer B.

I think cost analysis is critical, but that should all be done postmarketing because we as providers need to understand if this drug is 10

Page 326

times more expensive but it's going to save us some turnover time, that might still be a pretty significant run for the money to use the shorter acting, quicker agent, or if the actual number of rare but catastrophic adverse events was reduced by 50 percent.

But that's what regulatory's responsibility is, to take a look at that data, not look at the pharmacoeconomic data. That may change in the current political climate, but to me, unless I'm misrepresenting what I was taught as an advisor to the FDA, that's your job.

DR. WARD: Jenifer?

DR. LIGHTDALE: I'm just going to return quickly to two points actually. One is the blinding, and as the pediatric gastroenterologist in the room, I'll point out that we're talking about procedural sedation. So certainly, there are cases where the proceduralist is actually administering the sedation, which we can talk about, but there's also many of these situations you have a person administering the sedation and

Page 327

you have a proceduralist. You have a whole other person there who can be totally blinded. You want to remember that, and think about that creatively, I think.

I'm really intrigued by this issue of the cautious investigator, the person who -- sedation is fascinating, right? You're not giving a drug to somebody and they walk out the door, and maybe that adverse event happens while they're at home. The adverse event is happening right there in front of a provider trained to rescue them, and perhaps to try to keep them at some equilibrium so their heart rate doesn't go up and their blood pressure doesn't go down.

We have to capture what we're doing, or we're going to miss the fact that somebody's working really hard to create the noninferiority, if I articulated that.

DR. SESSLER: Excellent point.

DR. LIGHTDALE: Thanks, Dan.

(Laughter.)

DR. WARD: Comments?

Page 328

DR. WUNSCH: I just think that goes back to the ongoing theme today of this issue of different providers who will intervene at different times in different ways, and some who may sit back and let the O2 sat sit at 88 percent for an hour, if they're comfortable with it, versus someone else who's fussing when the sat starts to drop a little bit. So it's returning to the theme earlier in the day of that issue of recognizing different intervention thresholds and also just the amount of work that goes into a sedation as you're pointing out.

DR. WARD: Along that same line, the question also that I had was how should high-risk patients be incorporated into clinical trials, and that also has to do a little bit with the nervous observer. If I've got a patient that I don't think I can intubate and I'm doing a procedural sedation, I'm going to do it a little differently than if I've got a young healthy person that if they stop breathing, no big deal. I can ventilate them. I can intubate them. I can rescue them okay.

Page 329

A sleep apnea patient, as Ron pointed out, a pediatric patient with big tonsils, should we go out of our way to make sure that the clinical trials include patients that are at high risk?

DR. SESSLER: I'll be glad to comment. I guess this is part of the general discussion of internal validity versus generalizability. Clinical trials generally have good internal validity, which means that if you repeat the trial, you expect to get the same result.

One way that you control that is by minimizing variability. Investigators want to minimize variability anyway because for continuous outcome, the determinants of sample size are baseline variability and treatment effect. Treatment effect is the function of biology. Variability, you can control by whom you enroll in the trial.

If you look from a sponsor's perspective, a maker of a drug or of a device, for example, wants to have people in the study who are most likely to benefit and least likely to be harmed, and are

Page 330

relatively similar so that they don't impose much variability. That maximizes your chances of getting a statistically significant result with a manageable number of patients.

Those results are then taken by clinicians broadly and extrapolated to the whole world, and that's where you get into a problem because people take results from a highly selective clinical trial where not only the patients were selected, but the procedures were very well controlled and extrapolate that.

Experience has shown that these things actually don't extrapolate very well. In the real world, drugs are less effective and more toxic than they are in the original clinical trials, and the reason mostly has to do with selection.

Your point's really important. If you don't have representatives of the entire relevant population in your trial, you will get a result that does not apply to the entire population.

DR. WARD: Yes?

DR. SEXTON: It would seem to me, though, if

Page 331

the drug is not approved, it's really not an appropriate time, that the risk-benefit ratio for someone with a compromised airway who's an ASA class 3 or 4, before you know much about the drug is not the time to give it to that individual. So I would say don't enroll them in your study.

DR. WARD: So I guess my question would be how much do you need to know about the drug. And in these kind of drugs we're talking about, the efficacy is usually pretty straightforward. Either they provide sedation, or they don't provide sedation.

If you've got a drug that you've got enough so you know it works, it provides sedation, is that a pretty early point that you want to move right ahead and let's look at the sleep apnea patients, let's look with the kids with the big tonsils, let's look at the patients who are at higher risk.

At what point should that occur?

DR. SESSLER: That's why studies are phased, so in the initial phase, you're quite careful about whom you put into a trial, but by the time you're

Page 332

getting to phase 3 and you're trying to convince the FDA, and more importantly perhaps clinicians broadly, that this is an effective and safe drug, then you really should try to include the relevant population.

Too often, the relevant population is not included. It's in fact a subset of people most likely to benefit and least likely to be harmed.

DR. WUNSCH: I would think you would want to be maybe at the point where you know something about your risk profile of the drug beyond just whether or not it can create sedation.

For example, if you're dealing with a patient who's marginal in terms of their airway, and it turns out it is a drug where the risk profile errs on the side of more difficulty with airways, then obviously, that may not be the group that you then go next to, to assess. Similarly, if it's a drug that's shown a fair number of desaturations, your COPD patient who's already satting only 88 percent may not be the patients you're enrolling.

Page 333

So I think there probably would have to be some matching of at-risk patients and what is already known, and there would need to be enough known, about a new drug to feel confident that you're not compromising marginal patients, which probably puts you maybe one step beyond is it effective at causing sedation.

DR. WARD: With the caveat of what Dan just said, generalizability is not all that great. It may look as like a nice, safe drug as far as the airway is concerned in people with normal airways and be horrible in a patient with a -- and you wouldn't have any idea that that was going to happen because of the generalizability.

DR. SEXTON: That's probably one of the reasons the FDA requires postmarketing studies. That just seems like an appropriate time. Now, granted, I'm speaking from the point of the medical monitor, so I don't really prefer to have your patient be at tremendous risk and have to deal with that adverse event. That frightens me. Or it's not safe for the patient, or it doesn't seem like a

Page 334

good risk-benefit.

At least postmarketing, we're talking about the rule of three before. So if you have a rare fatal event, you've got to give that drug to an awful lot of people before you see it. So that would make me err on the side of caution in the clinical trial.

DR. WARD: Randy?

DR. CLARK: There's an aspect of this generalizability that we touched on this morning when we were talking about sedation in dental patients. Most of these trials are done in our acute care hospitals in the United States. The federal Medicare conditions of participation create a regulatory floor that all of these hospitals have to work under, and they're very specific about who is administering these drugs in what context, what the preoperative preparation is, interoperative monitoring, and post-anesthesia evaluation is.

There's no similar construct for what takes place in either physician offices or in dental offices, and those frequently then go to the states

Page 335

where they may be regulated by different boards.

For example, in Colorado, we have a sunset review portion of all of our practices boards, and whenever the dentistry boards come back up for sunset review, we in anesthesiology are always asked do you want to get into the issue of how dentists are providing sedation and anesthesia.

It gets to be a very politically complex issue, but I think we base a lot of the work that is done in dental offices on these clinical trials that are done in a very different environment with very different people.

I know that it was mentioned in California because of one of the recent complications there. A patient bill of rights is being looked at that would require in dental offices the same standards of care that might be required in an acute care hospital.

DR. WARD: Any other comments on this kind of area?

(No response.)

DR. WARD: I think one issue is if we're

Page 336

talking about rescue versus failure to rescue, because a lot of the complications are failure to rescue complications. In a well-designed clinical trial like Randy was saying, even high-risk patients in a well-controlled situation in which you could measure the need to rescue, would really give you a measure of, boy, we need to rescue a high percentage of patients who had airway obstruction.

That would give us a signal that if it's being used in a less well-regarded situation, in which rescue might not occur, that that would be a much more dangerous area.

I agree with you. There's a trade-off. At what point in what you know about the efficacy, do you then start aiming for the adverse event trials.

So along that line, my third question is -- because we heard about event-driven clinical trials to look at adverse event. What's the role of event-driven clinical trials in sedation to look at outcomes? The panel or anybody? I don't see those very often.

Page 337

DR. DEXTER: The study, which was done for pediatric radiology in Toronto, was very good in terms of the type of provider. Just the basic ideas week by week, and that's very standard or every other day, the study which was done for pediatric sedation in Finland, where they alternated every other day.

DR. WARD: Did they look for events? Was the size of the trial designed to look at occurrence of events?

DR. DEXTER: I don't think so. I think it would be inadequate for events. I think it was designed, both of them, in terms of time.

DR. WARD: Is there a role for event-driven trials for adverse events?

DR. SESSLER: Since the incidence is unknown when you start, you could end up with a pretty big trial if you're not careful. I'm not sure that there's a big role for it here.

All survival curve analyses are event driven. So whenever you have a trial where the outcome is, say, cancer recurrence, the sample size

Page 338

is not determined by the number of patients enrolled. It's determined by the number of outcome events, that is, the recurrences that happen. There certainly are lots of event-driven studies out there. I'm just not sure I see a huge role for it here.

DR. CRAVERO: It is really difficult, particularly in the pediatric realm because even at Boston Children's Hospital, where we have a high-risk population -- I'm almost surprised when I see somebody who actually has four chambers in their heart.

(Laughter.)

DR. CRAVERO: But even in that particular setting, when we try to come up with a risk stratification construct for our patients, we can select our certain characteristics that we know statistically make somebody more risky. But even putting that together, we come up with risk groups whose absolute risk is still relatively low, and you would need a great big trial.

Even of kids with Fontan physiology having

Page 339

sedation for their MRI or whatever, the actual number of events, although it's orders of magnitude higher than it would be for other people, still relatively low.

So I would agree, it's really hard to do that kind of a study at least in the pediatric population. I think in the adult population where there's a lot more comorbidity and perhaps more events to look at, you're talking about a different situation.

DR. WARD: Other questions from -- we've had a busy day. I don't mind getting through 15 minutes early. Is everybody all --

DR. CLARK: I'll just respond to Joe's comment about ventricles. At Denver, we have an active adult congenital cardiac disease program where cardiologist cross both sides of the street between University Hospital and Children's, we have the two ventricle patients taken care of at University Hospital, and the one or fewer ventricle patients at Children's.

(Laughter.)

Page 340

Adjournment

DR. WARD: Thank you all. Dinner tonight at 7:00. I think it's been a very successful, productive day. I look forward to the second one tomorrow. So we'll see you-all at dinner.

(Applause.)

(Whereupon, at 4:24 p.m., the meeting was concluded.)

B

C

D

E

F

G

H

I

M

N

O

Q

S

U