DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

+ + + + +

RISK EVALUATION AND MITIGATION STRATEGY (REMS) ASSESSMENTS: SOCIAL SCIENCE METHODOLOGIES TO ASSESS GOALS RELATED TO KNOWLEDGE

+ + + + +

PUBLIC WORKSHOP

DOCKET NO. FDA-2012-N-0408

+ + + + +

THURSDAY
JUNE 7, 2012

+ + + + +

The Public Workshop met in Building 31, Room 1503 at the Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, Maryland, at 8:00 a.m., Theresa Toigo, Moderator, presiding.

PRESENT

THERESA TOIGO, RPh, MBA, Associate Director for Drug Safety Operations, CDER; Moderator

CLAUDIA MANZO, PharmD, Director, Division of Risk Management, FDA Office of Surveillance and Epidemiology (OSE)

MARY WILLY, PhD, Associate Director, Division of Risk Management, OSE

ROBERT SHIBUYA, MD, Deputy Director, Division of Risk Management, OSE

PRESENT(Cont'd)

PAUL BEATTY, PhD, National Center for Health Statistics, Centers for Disease Control and Prevention; Panel Member (Panel 1)

DAVID CANTOR, PhD, Westat; University of Maryland; Panel Member (Panel 1)

SADEQ CHOWDHURY, PhD, Agency for Healthcare Research and Quality; Panel Member (Panel 1)

TERRY DAVIS, PhD, Louisiana State University Health Sciences Center-Shreveport; Panel Member (Panel 2)

RUTH DAY, PhD, Duke University; Panel Member (Panel 2)

SARA EGGERS, Office of Planning and Informatics, CDER; Panel Member (Panel 2)

JONAS ELLENBERG, PhD, Perelman School of Medicine, University of Pennsylvania, Panel Member (Panel 1)

ROBERT FAY, PhD, Westat; University of Maryland; Panel Member (Panel 1)

ALICIA GILSENAN, PhD, RPh, RTI Health Solutions; Panel Member (Panels 1-2)

SANJAY GUPTA, PhD, MBA, Daiichi Sankyo, Inc.; Panel Member (Panels 1-2)

WILLIAM HOLMES, MD, MSCE, AstraZeneca Pharmaceuticals; Panel Member (Panels 1-2)

KENNETH HORNBUCKLE, DVM, MPH, PhD, Eli Lilly and Company; Panel Member (Panels 1-2)

KAROL KROTKI, PhD, RTI International; Panel Member (Panels 1-2)

LISA LAVANGE, PhD, Office of Translational Sciences, Food and Drug Administration; Panel Member (Panel 1)

MATTHEW LEE, PharmD, Lundbeck, Inc.; Panel Member (Panels 1-2)

SUNGHEE LEE, PhD, Institute for Social Research, University of Michigan; Panel Member (Panel 1)

PRESENT(Cont'd)

DORIS LEFKOWITZ, PhD, Agency for Healthcare Research and Quality; Panel Member (Panels 1-2)

STANLEY LEMESHOW, PhD, School of Public Health, Ohio State University; Panel Member (Panels 1-2)

ELAINE MORATTO, DrPH, MPH, Colorado School of Public Health, University of Colorado; Panel Member (Panel 2)

LOUIS MORRIS, PhD, Louis A. Morris and Associates; Panel Member (Panels 1-2)

NANCY OSTROVE, PhD, Office of the Commissioner, Food and Drug Administration (retired); Panel Member (Panels 1-2)

JANET PETERSON, PhD, Janssen Scientific Affairs; Panel Member (Panels 1-2)

TRIVELLORE RAGHUNATHAN, PhD, Institute for Social Research, University of Michigan; Joint Program in Survey Methodology, University of Maryland; Panel Member (Panel 1)

SAUL SHIFFMAN, PhD, University of Pittsburgh; Pinney Associates; Panel Member (Panels 1-2)

BETSY LYNN SLEATH, PhD, University of North Carolina; Panel Member (Panel 1)

MEREDITH SMITH, PhD, MPA, Abbott Laboratories; Panel Member (Panels 1-2)

JOLENE SMYTH, PhD, University of Nebraska-Lincoln; Panel Member (Panel 1)

ANNETTE STEMHAGEN, DrPH, FISPE, United BioSource Corporation; Panel Member (Panels 1-2)

GORDON WILLIS, PhD, National Cancer Institute, National Institutes of Health; Panel Member (Panel 1)

MICHAEL WOLF, PhD, MPH, Feinberg School of Medicine, Northwestern University; Panel Member (Panel 2)

ALSO PRESENT

DORIS AUTH, PharmD, Pharmacist, Food and Drug Administration

BRIAN GORDON, MA, Social Science Analyst, Food and Drug Administration

RIMA IZEM, PhD, Math Statistician (Biomed), Food and Drug Administration

CONTENTS

Welcome, Introductions and Logistics.....7
Moderator, Theresa Toigo, R.Ph., M.B.A.
Associate Director for Drug Safety Operations, CDER

Opening Remarks.....12
Claudia Manzo, Pharm.D. Director, Division of Risk Management, FDA Office of Surveillance and Epidemiology (OSE)

FDA Presentation: Introduction to REMS and REMS Assessments.....16
Mary Willy, Ph.D., Associate Director, Division of Risk Management, OSE

Panel 1 Introduction and Charge.....24
Robert Shibuya, M.D., Deputy Director, Division of Risk Management, OSE

Industry Experience in Using Surveys to Assess REMS' Impact on Knowledge, Attitudes & Behavior.....40
William C. Holmes, M.D., Medical Director, AstraZeneca Pharmaceuticals
Meredith Smith, Ph.D., M.P.A
Senior Scientific Director, Risk Management, Abbott Laboratories

Open Public Speakers Session.....70

Panel Discussion: Consideration for primary endpoints.....107

Panel Discussion: Strategies for recruitment and sample size considerations.....171

Panel Discussion: Questionnaire design....218

Panel Discussion: Process issues.....269

CONTENTS(Cont'd)

Panel 2 Introduction and Charge.....302
Robert Shibuya, M.D., Deputy Director, Division of Risk Management, OSE

Open Public Speakers Session.....316

Panel Discussion: Alternative to surveys..331

Panel Discussion: Assessing behavior, burden and access 386

Summary of discussion.....416
Robert Shibuya, M.D., Deputy Director, Division of Risk Management, OSE

Conclusion.....418

7

P-R-O-C-E-E-D-I-N-G-S

(7:58 a.m.)

MODERATOR TOIGO: Good morning. My name is Terry Toigo and I serve as CDER's associate director for drug safety operations. I welcome you to this public workshop to discuss risk evaluation and mitigation strategies, or REMS, assessments: social science methodologies to assess goals related to knowledge.

As you can see from the agenda, we have a very full day. We have two very large panels to address six challenging topics. We have 15 speakers in our open public session, and we have a few presentations by FDA and industry representatives to provide some context for our discussions today.

Dr. Shibuya will introduce our panel members at the start of each of the two panels and our speakers will be available to answer clarifying questions at the beginning of the panel sessions, as opposed to after

their presentations.

A few housekeeping details before I turn the podium over to Dr. Claudia Manzo, our first speaker.

For the record, we didn't receive any requests for accommodations due to disability for this meeting, so that's taken care of.

I ask everyone to silence their cell phones, your BlackBerrys or other devices if you haven't already done so.

For our panel members, please remember to turn on the microphones and speak into the microphones every time you are recognized to speak, and then turn them off when you are not speaking. This will take a while for us to get used to, but that's the way it works.

We are transcribing the meeting so it would be helpful to the transcriber and also to members of the audience if you state your name before you begin to speak.

We ask that all of the attendees sign in at the registration table at the front. Open public session speakers need to sign in at the speaker registration table, and if you have not checked in at the meeting registration desk and you are signed in to speak, please do so. That way we can make sure we have everybody and we have the timing, and we get through -- smoothly through the open session. And alternatively, if you choose not to speak but you still want to make sure your comments are heard, we have the docket open until July 7.

Agendas, if you didn't already get them, are available at the registration tables, and all of the other materials will be available on the FDA website, the meeting link, you can get the other materials from the meeting.

Karen Mahoney from CDER's office of communication will be the press contact if there's anyone interested, and Karen is in

the back waving her hand over there. So if you need Karen she is there for you.

The slide presentations for all the speakers will be posted on the FDA website with the meeting announcement. Transcripts of the meeting will be available about 30 days after the meeting, and details on how to access the transcript are available at the bottom of the agenda.

Again, the docket is open to July 7 for feedback, and that includes panel members. We have a lot of people and we recognize that we are not going to get everybody, everything they want to say, but we think it's important that we hear what you have got to say, and that's the purpose of the docket, that the docket comments are publicly available, and we review those comments, and we use them, so it's not just this meeting, it is the docket. So, panel members, if you don't get a chance to say something really important, that docket is

available for you.

Now, the housekeeping. The restrooms are available outside the conference room in the back of the lobby. You go to the exit up the hallway. Before your exit, you turn right and then you go straight to the end of the hallway and they are on either side.

We will have a 15-minute break around 9:30, and then in the afternoon, around 2:00. Lunch will be at 12:30, hopefully I'll keep things on time. Coffee and tea will be available for sale during the breaks, and lunch will also be available here. You can leave the building, but chances are you are not going to get back in an hour.

And a final housekeeping detail, your feedback about this meeting is really important to us, and that's the positive and the negative feedback, because what you tell us will help us plan future meetings. So you

can see me, or you can see any one of the FDA staff, or you can email me after the meeting as to what you think went well and what didn't, and that's theresa.toigo@fda.hhs.gov.

So we welcome your feedback. And now we will begin our first presentation. Claudia Manzo is our first speaker. She is the Director at the Division of Risk Management in the Office of Surveillance and Epidemiology and the Center for Drug Evaluation and Research, and she is going to give us some brief comments to kind of provide context for the day's discussion. So, Claudia, welcome.

DIRECTOR MANZO: Good morning and welcome to all the panel members and the audience. Since the approval of REMS in 2007, FDA has been interested in obtaining feedback from our stakeholders about the implementation and the evaluation of REMS.

In July 2010, FDA held a meeting to obtain input on issues and challenges

associated with the development and implementation of REMS.

One of those sessions focused on the evaluation of REMS and the use of surveys to assess patient and healthcare provider understanding of the serious risks of a drug, and how to use the drug safely.

At that meeting, we heard that surveys are considered the standard method to assess knowledge, but do not necessarily inform behavior.

Since the 2010 public meeting, FDA has gained extensive experience reviewing survey methodologies and instruments, as well as the results of surveys that have been conducted as part of an assessment of a REMS.

We are interested in sharing our experience and continuing the discussion about the use of surveys as an assessment tool. In addition, we would like to discuss other options for assessing knowledge.

There are five objectives of this

meeting. The first is to initiate dialogue and information sharing about survey methodologies and instruments used to evaluate patients' and healthcare providers' knowledge about the risks of certain drugs.

Next, we would like to share our current experience regarding surveys as a component of the REMS assessment plans. We want to obtain information that could be used to develop standard methods for evaluating patient and provider knowledge, and fourth, to discuss alternative methodologies to assess knowledge.

And finally we would like to discuss the use of surveys as a tool to assess patient and provider behavior changes, burden on the healthcare system and patient access to the drug under a REMS.

Today's panelists including the following experts: social science professionals, particularly those with survey design and analysis expertise; health

communication and health literacy experts; statisticians; pharmaceutical industry representatives; and representatives from other government agencies that conduct national surveys.

The input we receive from the panel may be used to develop guidance for industry describing best practices for conducting an assessment of patient and provider knowledge about a drug's risks.

This morning, during panel 1, there will be two FDA presentations. Dr. Mary Willy will present an overview of REMS, and Dr. Robert Shibuya will present a summary of FDA's experience with surveys and assessment tools, including identifying challenges.

Drs. William Holmes and Meredith Smith will follow with presentations that summarize industry response and comments on survey methodologies. We will spend the next few hours discussing a series of survey-

related questions.

In the afternoon, panel 2 will discuss -- there will be a short introduction by Dr. Shibuya and then a panel discussion about alternatives to surveys and options for utilizing surveys to assess other REMS-related information, such as burden on the healthcare system and on patient access. Thank you.

MODERATOR TOIGO: Thank you, Claudia. And now Dr. Mary Willy will give a brief presentation on REMS assessments and Dr. Willy is associate director of the Division of Risk Management in the Office of Surveillance and Epidemiology in CDER. Welcome, Dr. Willy.

ASSOCIATE DIRECTOR WILLY: Good morning. As part of the introduction to this workshop, I am going to provide a very short overview of the FDA's current risk management authorities.

And just as a little background,

the risk management of drugs at FDA is not a new effort. The REMS provision of the 2007 Food and Drug Administration Amendments Act, which is often referred to as FDAAA, were based on the 1992 restricted distribution regulation, and the 2005 guidance for industry on the development and use of risk minimization action plans.

In March 2008, FDAAA took effect, and it gave FDA a new series of authorities, including the ability to require sponsors to develop and comply with risk evaluation and mitigation strategies, also called REMS.

A REMS is a required risk management plan, as specified under FDAAA, and it goes beyond routine professional labeling to ensure the benefits of a drug outweigh the risks.

The FDA can require a REMS at the time that drug is approved, and a REMS can also be required after approval if there is new safety information. And it's good to

remember that the FDA's authority is over the sponsor, and not the healthcare professionals.

So the REMS can include a number of elements and I will describe them in more detail shortly, but just to give you the list, they can include: a Medication Guide or Patient Package Insert; a communication plan; elements to assure safe use, what we often refer to as ETASU; an implementation system; and a timetable for submission of assessments.

The FDA determines the final elements that are going to be part of a REMS after considering the magnitude of the risk and the tools that might be used to manage that risk.

So I'll go over these different elements. The Medication Guide is a patient-friendly instrument that is distributed with the drug. After passage of FDAAA, drugs that were approved with a

Medication Guide were required to have a REMS. In November of 2011, the FDA published a final guidance and that guidance limited the number of REMS that would require a Medication Guide as we moved forward.

A communication plan is another REMS element, and communication plans include drug information that will be used by healthcare providers. Communication plans usually include a Dear Healthcare Professional letter, but they can also include information for professional organizations or societies.

The elements to assure safe use, there can be a number of them in this section and I'll go through them briefly. First of all, there can be a requirement that healthcare providers who prescribe the drug have certain training or that they are specially certified.

There can be a requirement that pharmacies and practitioners or healthcare

settings that dispense the drug are also certified.

The drug may be required to be dispensed in certain healthcare settings, like hospitals. The drug may be dispensed to patients with evidence of safe-use conditions, such as a certain enrollment or the patient signing a certain form.

There may be a requirement that patients undergo certain monitoring, such as liver enzyme monitoring or pregnancy testing, and there may be a requirement that the patients enroll in a registry.

An implementation system may be required if there are certain elements in the REMS, particularly if there is a requirement for certification of pharmacies and hospitals, if there's a limitation to the use of certain healthcare settings, if there's a requirement for safe use conditions.

The implementation system is put in place to monitor that these elements are

actually implemented, and if there are problems to develop strategies to improve.

So, when a drug is approved, there is an expectation, a requirement that an assessment is done, and it's usually -- and it's submitted at 18 months, three years and seven years, but it can be more frequent.

The details of the assessments are included in the approval letter. The assessment plan that describes the methods and the rationale for the different metrics is included in the supporting document.

So, an assessment plan, when there's a Medication Guide, will include the assessment that evaluates the patients' understanding of serious risks.

Since the Medication Guide can include a fair amount of information, we recommend that the sponsors focus on the section of the Medication Guide that is titled, "What is the most important information I should know?"

The sponsors are asked to submit the proposed methodology for the surveys, and they can submit them at any time, but generally we ask that they submit them at least 90 days before conducting the evaluation so that the social scientists in our office can review them.

Since the communication plan can include education for prescribers, assessments of REMS with communication plans usually includes an evaluation of those efforts.

And the assessment of a REMS that includes elements to assure safe use will include many more metrics, because they are more complicated REMS, and often a good part of the assessment is metrics about the process.

But in addition, because there's education involved, evaluation of knowledge of the patients and/or prescribers is part of that assessment.

This is a very busy slide and it just summarizes the activity we have had around REMS since approval in 2008. So we have approved 200 individual REMS, 125 of those REMS were what we called Med Guide-only and that means that they were REMS that only had a Medication Guide and no other element.

Because of that 2001 guidance that I have referred to, 105 of those REMS have been released. And so at this time we have 86 approved REMS that are currently active, and nearly half of those are REMS that have elements to assure safe use.

I do want to note that, for those of you who are not familiar with the word or the term "Deemed REMS," those are programs that were actually in place before FDAAA and are being transferred into the new system. We also have two single shared systems, and these are programs that include multiple sponsors.

And that's my summary. I hope you

find it helpful. Thank you.

MODERATOR TOIGO: Thank you, Mary. For our final scheduled presentation, we are going to hear from two representatives from the -- actually, Rob is going to go before them. Rob is going to provide the overview of both of the panels for today, starting just with panel 1 this morning, panel 2 this afternoon.

DEPUTY DIRECTOR SHIBUYA: Good morning and welcome to the FDA. I'd like to start by asking the panelists -- I'd like to actually start by thanking the panelists for their willingness to participate this morning, and I'd like to start by asking the panelists to introduce themselves, maybe starting at my 12 o'clock.

In the interests of time, please limit yourself to just your name and your institution.

PANEL MEMBER BEATTY: Paul Beatty, National Center for Health Statistics.

PANEL MEMBER CANTOR: David Cantor, Westat and JPSM, Joint Program for Survey Methodology.

PANEL MEMBER CHOWDHURY: Sadeq Chowdhury from Agency for Healthcare Research and Quality.

PANEL MEMBER ELLENBERG: Good morning, Jonas Ellenberg from the University of Pennsylvania.

PANEL MEMBER FAY: Bob Fay, Westat and JPSM, University of Maryland and Michigan.

PANEL MEMBER GILSENAN: Hello, Alicia Gilsenan from RTI Health Solutions.

PANEL MEMBER GUPTA: Hi, Sanjay Gupta from Daiichi Sankyo.

PANEL MEMBER HOLMES: William Holmes from AstraZeneca Pharmaceuticals.

PANEL MEMBER HORNBUCKLE: Good morning. Ken Hornbuckle from Eli Lilly and Company.

PANEL MEMBER LAVANGE: Lisa

LaVange, director of Office of Biostatistics in CDER.

PANEL MEMBER MATTHEW LEE: Good morning. Matthew Lee from Lundbeck.

DR. IZEM: Good morning. Rima Izem from the Office of Biostatistics at CDER.

MR. GORDON: Brian Gordon from Division of Risk Management.

ASSOCIATE DIRECTOR WILLY: Mary Willy.

DIRECTOR MANZO: Claudia Manzo, Division of Risk Management.

MODERATOR TOIGO: Terry Toigo, associate director for Drug Safety Operations in CDER.

DR. AUTH: Doris Auth, Division of Risk Management.

PANEL MEMBER WILLIS: Gordon Willis, National Cancer Institute.

PANEL MEMBER STEMHAGEN: Annette Stemhagen, United BioSource Corporation.

27

PANEL MEMBER SMYTH: Jolene Smyth, University of Nebraska-Lincoln's survey research and methodology program.

PANEL MEMBER SMITH: Meredith Smith, Abbott Laboratories.

PANEL MEMBER SLEATH: Betsy Sleath, University of North Carolina, Chapel Hill.

PANEL MEMBER SHIFFMAN: Saul Shiffman, University of Pittsburgh and Pinney Associates.

PANEL MEMBER RAGHUNATHAN: Trivellore Raghunathan, University of Michigan and also at Joint Program in Survey Methodology at the University of Maryland.

PANEL MEMBER PETERSON: Janet Peterson, Janssen Pharmaceuticals.

PANEL MEMBER OSTROVE: Nancy Ostrove, ex-risk communication staff, FDA, currently myself.

PANEL MEMBER MORRIS: Lou Morris, LAMA.

PANEL MEMBER LEMESHOW: Stan Lemeshow, Ohio State University.

PANEL MEMBER SUNGHEE LEE: Sunghee Lee, University of Michigan and also at JPSM.

DEPUTY DIRECTOR SHIBUYA: Okay, so again, good morning. In this presentation, I will provide an introduction to panel 1, entitled "Surveys to Assess Knowledge."

I'd like to make two key points to start. First, in this panel, we do presuppose that surveys are a reasonable method to assess knowledge in the context of REMS.

I note that we have not yet published guidance in this area. We have identified issues that I will describe shortly as we have received and reviewed surveys submitted as part of REMS assessments. We now wish to hear a discussion of how to address the issues identified from this expert panel.

I am going to cover the following

topics. I'll charge panel 1 with discussing certain issues related to these surveys. I'll show examples of language typically appearing in approval letters, the result of which is a survey of patients or prescribers.

I will then provide a qualitative description of what we have observed in the surveys submitted to date.

In describing what we have seen, I will expand on some observations that have led to questions and concerns about these surveys. Last, I'll be available to answer clarifying questions.

The goal of panel 1 is -- and I'll just read it -- to provide input regarding specific, feasible strategies to improve surveys that are submitted to assess the educational components of REMS.

We will ask you to discuss the following six items: survey endpoints, threshold for success, survey participant recruitment, sample size, question design and

process issues.

This slide contains deidentified but otherwise verbatim language from a REMS, and that's for the goals. And from an approval letter, and that's from the part that pertains to the requirement for the assessment for a certain -- a deidentified drug.

And as you can see, I'm just going to read the first goal, because the second reads identically. The goal is to inform healthcare providers about the risks associated with a certain drug.

And then the required assessments reads, "Evaluation of healthcare providers' understanding and patients' understanding of the risks of the drug."

The next part of the talk involves a summary of what we observed in our reviews of surveys. As you can see, we received our first assessment over three years ago in September of 2008, and as of a cutoff at the

end of December of 2011, we have received 144 assessments with surveys.

You can see about 55 were Medication Guide-only, and 56 were other, which were predominantly REMS that included ETASU, and many of those had serial assessments or multiple surveys on the same product, and then the balance of them were a combination of Medication Guide and communication REMS.

The next two slides summarize qualitative observations we have made about the surveys we have received to-date. We will be asking you to discuss each of the topics on this slide.

So to start out with, we note that in general, the surveys do not contain pre-specified -- any sort of rigorous, pre-specified statistical analysis, or there is no formal hypothesis testing.

Most of these surveys are not piloted or pre-tested. The surveys almost

universally use convenience samples that we define as really what's most feasible for the investigator. The most common prescriber recruitment strategy is to use a pharmacy -- typically a pharmacy service or an audit service to identify prescribers and then to mail the letter inviting them to participate in the survey.

With regard to recruiting patients, sponsors typically, again, use some sort of pharmacy system to identify patients who could be eligible for the survey, or when it's not -- when it doesn't go -- when the drug is not distributed through a pharmacy, sometimes they'll ask the healthcare provider to identify and invite the patient.

This slide contains a listing of information generally collected in surveys. You'll not specifically be asked about these items, but the data collected include obviously questions about the knowledge of risks, patient or prescriber demographics,

how long the patient has been on the drug or how long the prescriber has been prescribing drug.

And then there are required metrics pertaining to Medication Guide distribution. Does the patient remember getting the guide, where did they get it, did somebody offer to explain the Medication Guide to them.

You are going to be asked -- this slide contains a description of the endpoint normally used in these surveys. This will also be a discussion question.

And so the endpoint most commonly used we call the knowledge rate, although we acknowledge there is not a time component of this. But the typical endpoint is basically the results -- the number of respondents who answered correctly over the number of respondents who answered the question, and that's expressed as a percentage.

I'll also note that some surveys

use a composite scoring system, where they ask a risk concept in multiple different ways, and in order to be considered a successful sort of responder you need to typically answer three out of four or four out of four, or some proportion over 50 percent correct.

You are going to be asked to discuss thresholds from which one could conclude that the desired knowledge was adequate in the population.

So we have observed that the general definition, to paraphrase it, is sort of here: it's the proportion of respondents responding correctly, which, if met or exceeded, would support a conclusion that the desired knowledge was adequate.

The minority of surveys do specify a threshold, but I will say that it's pretty consistent that when they do, 80 percent is the normal threshold selected.

An important issue that we'll ask

you to discuss regards sample representativeness. We have some evidence that the person surveyed may not be representative of the population of interest.

In 2010, we conducted a small pilot study to assess this question. As this slide indicates, the limited amount of demographic information it surveys was compared to patient and prescriber information from the Physician Drug and Diagnosis Audit and differences in the characteristics were assessed.

Briefly, again, this was a small pilot study, and consisted of data from seven patient surveys and two HCP surveys. Five of the patient surveys and both HCP surveys showed differences between the population and PDDA data.

These data must be interpreted with extreme caution. As noted here, the survey had -- the study had a number of limitations, and I've noted them here.

At the time the study was conducted, the FDAAA had been in place for a very short time and we really had a very small number of surveys to work with.

Many of the metrics of interest to look at generalizability are really not captured. Those are things like income, education, and health literacy.

The demographic patterns for products with low use may not be generalizable, and, again, the level of detail of the demographic information collected was variable.

Another topic for discussion will be sample size. As this slide shows, some surveys have completed very few participants, as few as eight for patients and four for prescribers.

As shown in the slide, the reason usually given for low sample size is small patient and prescriber populations. I will note, though, that not infrequently the

methodology will specify a target sample size, usually around 300 or 400.

Sometimes that target sample size is not met. Most methodologies use the following rationale to estimate sample size, and since I a not a statistician, I will just read it.

It says it's based on the precision of knowledge rate estimate. Common assumptions are: margin of error of six percent, confidence level of 95 percent and knowledge rate of 50 percent.

We'll also be asking you to comment on the question type used. Generally, sponsors use the true, false and I don't know question form. That's, I think, in the vast majority of questions.

The other used question form is multiple choice, both the single best answer and -- or choose the best one amongst -- or choose all applicable answers.

And I'll note that it's really a

very small number of surveys have used clinical vignettes as part of a question.

I have a few other observations to make prior to opening it for clarifying questions. First, there is the obvious potential for confounders, such as the fact that -- confounders, such as the level of difficulty of questions, the lag time -- frequently it's not particularly known what the lag time was between when the individual was messaged and when they were surveyed, also sort of the difficulty of the concepts to be conveyed.

Again, I'll note that not all surveys are pre-tested. It really seems to be the minority are, which leads to difficulty interpreting the results at times.

And then last, as I said there, because healthcare providers and patients obtain information from different sources, the contribution of REMS cannot be clearly determined.

And with that, I -- Terry, I don't know if you want to wait for questions until after the industry presentation or --

MODERATOR TOIGO: I thought we were going to do the clarifying questions at the beginning of the panel, unless there is a question that really needs to be -- Rob needs to address now. So it looks like Dr. Ellenberg.

PANEL MEMBER ELLENBERG: In reviewing the materials that you sent, in terms of what the REMS look like for each class or each individual drug, and in listening to your comments just now, one of the questions, clarifying questions, that arises, at least for me, is whether or not you are looking for advice in terms of assessing an approved, single REMS document, and how it's actually heard and listened to, or are you looking for advice in terms of, not only the questions you have raised, but in terms of looking at alternatives to the

document that is the official REMS, looking at alternative ways of asking the questions?

For example, do you want to get advice for us on whether a five-question survey to assess the REMS document as a communication tool is better than a 10-question survey?

DEPUTY DIRECTOR SHIBUYA: So I think you -- there are several questions in there. The first question is the -- you talked about alternatives, and that really is the subject of panel 2.

So there's a lot to cover, so that's why we kind of would like to make a clear distinction between panel 1 and panel 2. So for panel 1, we are really looking for -- I know we sent you the examples, but that was just to give you a sense of what we are seeing. Generalizable -- I would say generalizable comments about how to improve these surveys and how to address the issues that we have identified is really what we are

looking for from you today, this morning. Does that answer the question?

PANEL MEMBER ELLENBERG: Yes, it does. Thank you.

MODERATOR TOIGO: Looks like that's it for the clarifying questions, Rob.

Okay, so our list presenters for this morning or for this session will be Dr. William Holmes from AstraZeneca and Dr. Meredith Smith from Abbott Laboratories, and they are going to talk about the industry experience in using surveys to date.

And then, as I mentioned, we will have other industry perspectives that will be provided during the open public session later. Dr. Holmes.

PANEL MEMBER HOLMES: Good morning and thank you. I am William Holmes from AstraZeneca and it's a pleasure to be here today. Dr. Meredith Smith and I will be presenting on behalf of those listed on the bottom of the slide here, in which we are

going to be discussing industry experience in using surveys to assess REMS impact on knowledge, and in particular, describing some learnings we have drawn from that experience.

On behalf of all of us on the panel here today representing industry, I would like to thank the FDA for arranging this public workshop and for inviting us to participate as industry representatives on the panel.

My colleagues and I would also like to acknowledge FDA's response to stakeholder feedback through enforcing Medication Guides outside of REMS and for focusing efforts to improve communication to patients through patient-centered initiatives, such as the effort to standardize patient medication information.

We'd also like to make a disclaimer: specifically, that the content of this slide presentation represents the collective opinions and experiences of the

individuals chosen by PhRMA to serve on this panel. It does not represent the views of any particular organization or pharmaceutical company.

Over the next 20 or so minutes, we will start by briefly affirming the value of knowledge surveys, then we will follow with a number of slides discussing methodological challenges of using surveys to assess knowledge, including questions about whether knowledge is the only or most appropriate outcome of these assessments, concerns about the psychometric rigor of these assessments, as well as what other data collections might be used and some other research design issues, as well, before we make some concluding remarks, and then we will finish with other panel members representing industry making additional comments.

I am going to give you a few seconds to just sort of sit with this slide and look through it for a second.

44

We wanted to begin with this slide to highlight what most, if not all of us, in this room know about the underpinnings of public health interventions, that there are a number of theoretical models with supporting empirical data showing that a change in knowledge can lead to a change in attitudes, that in turn can lead to a change in behaviors.

Given that risk mitigation efforts are public health interventions with the ultimate goal of change in behaviors, knowledge and changes in knowledge are not the only goal about which we wish to obtain information.

In the end, we want to know that we have changed behavior, the risk mitigation itself. As that will always be difficult to assess, we must base assessment on theory, whereby all aspects, including mediators and moderators of the knowledge to behavior paradigm, are measured, including variables

such as attitudes, subjective norms, behavioral intent and self-efficacy, and there's a number of guidances by which this can be done, including those from the FDA that we have there on the slide.

At the same time, we want to stress our agreement with the importance, even if not the primacy, of knowledge assessments, and that's what this slide is about.

We concur that it's important to measure patients' and healthcare providers' knowledge related to drug safety risks, that it's valuable to obtain systematic feedback from end users, whether that be patients or healthcare providers, it's important to assess important to assess the effectiveness of existing approved methods for educating patients and HCPs using the methods that have been mentioned already this morning, and that self-report is an only way by which to obtain some of this information. So, for example,

knowledge about the risk and the risk mitigation on the parts of patients and HCPs.

And finally, as we stressed on the prior slide, there's many federal agencies that have substantial expertise in conducting surveys and have guidances related to those.

All of that said, there are methodological and other challenges related to these sorts of assessments. And the first of those that we'd like to talk about having some learnings about in the process of having done many of these across the industry representatives on the panel, the first and perhaps biggest is maybe the psychometric challenges associated with knowledge questionnaires.

We know that without reliability and validity that we don't know that we are measuring what we say we are measuring, nor how well we are measuring what we say we are measuring.

So in doing any knowledge

47

assessment, we need to establish the content validity of the questionnaire via qualitative and quantitative methods and these are well established both in the information I talked about in the prior slides in terms of guidances, and in other social science texts et cetera, there are qualitative approaches to generating questionnaires and then pre-testing of those questionnaires and approaches to doing the psychometric piece of establishing reliability and validity.

And I'm distilling a whole, you know, something that can take semesters to study and teach, but distilling that down into a few bullets is overly simplifying, but among other things is the need to develop and include multiple items measuring the same construct.

And the application of best psychometric practices can be extremely challenging with multiple stakeholders contributing input, and the one example we

wanted to provide here is that sometimes on industry side there is efforts put forward to construct questionnaires with some best psychometric practices and then it goes in for review, and then -- so we might have a group of items that are assessing a particular construct and then we get back in positions of either removing an item or adding an item or more, and we now have a group of items that had been previously tested somewhat and perhaps assessed a construct, with a new set of items that we don't know anymore because we don't have that qualitative work or we have qualitative work to suggest that some of those items actually worked well and they have now been removed, and they now are what become the tool by which we are assessing a construct that we had perhaps limited information about before, about its content validity, but certainly now no longer do.

I think the other point, and the

final bullet, is the questionnaire content and structure needs to be closely linked to REMS goals. So those best psychometric practices can sometimes be following that closely and then again, as we have multiple stakeholder input, sometimes there are in positions of other types of questions that may be in addition to those REMS goals, and then we have constructs again that are expanded or unclear and perhaps the intent to pursue best psychometric practices become challenged at the very least.

Okay. So the next two slides -- I'm sorry, I'm having a little trouble here with, there we go -- so the next two slides grapple with one of the other -- so the first very big learning that we would like to suggest is the psychometric challenge. The second big learning is grappling with one of the other substantial methodological challenges, which is just what are we measuring?

Are we measuring knowledge? Should we be measuring knowledge? There are many related constructs as well as mediators and moderators of knowledge that we may be measuring or should be measuring and these include -- and again, these are spread across two slides, and I am going to mostly read these bullets.

So for example, the first is exposure. Has the respondent been exposed to the risk education materials? Did they find it useful and acceptable as information? Could they navigate the -- for example, the patient -- could they navigate the Med Guide and obtain the information that we really wanted them to obtain?

Did they comprehend it? And across these two slides, knowledge is sort of right in the middle of this because there's lots of things preceding knowledge and lots of things after knowledge.

So even at knowledge level itself,

when we are looking at that and testing it, are we interested in whether they recognize something, or whether they can actually regurgitate it? Can they self-generate it?

So if you think about knowledge about MI symptoms, can you tell me what those symptoms are if I asked you what they are? Or is it more interesting for me to ask you to show you a bunch of symptoms and have you circle them? So that's the difference between self-generating versus recognition. So do we want regurgitation or do we just want recognition? Unclear.

Then we are -- a number of things here in terms of outcome of interest is, are we interested in retention of knowledge? How long are we interested in that retention of knowledge? Functional understanding. Can someone apply the abstract knowledge they have in a real-world scenario, and do they have the self-efficacy to do that? So that's a moderator between the comprehension to

behavior process.

Then there's behavioral intent, the ability to and intent to act on the acquired information and knowledge, and then finally the actual behavior. Can somebody perform the desired skills or behavior, can they report symptoms et cetera.

So across these two slides, you see a whole range of things that are, some of them, knowledge-related, some of them knowledge itself, and even then what sort of knowledge are we interested in, and then from the knowledge to behavior range, a number of things as well.

The next two slides touch on a number of research design issues, starting with the lack of baseline data. So if you think of some approaches, for example educational research in terms of interest in knowledge acquisition, you may have an intervention that provides the information that you are interested in having knowledge

gained, and you have a pre- and a post-test, so pre-test intervention and post-test. There is no sort of pre-test here of what sort of baseline knowledge or comprehension people have. There is no comparator arm.

And these two slides are a bit wieldy, covering lots of things. So follow with me I hope.

The next bullet is about evaluation intervals. So as we saw mentioned a few minutes ago, often they are 18 months, 3 and 7 years in terms of the assessment time periods.

There are others in which there is a six-month time period. There's others in which there's annual assessments. And so if you think back to one of the first research challenges that were mentioned, which is the best psychometric practice, those of you who do psychometric work in the room, know that these can be quite lengthy and entail a great deal of effort, and when you have an interval

of 12 months to do your REMS and then assess how the REMS is doing, to do good psychometric work that buttresses those assessment studies can be challenging to say the least.

And that's, let's say, in your first assessment period. If you have a second assessment period 12 months later, and you want to get FDA response, your FDA response can sometimes come after the cutoff by which you need to start your assessment period, and so one can't always even be responsive to what all the stakeholders may be wanting to indicate should be done.

There's challenges in terms of appropriate power. So if we are not clear about the primary endpoint, or the primary endpoint is challenging, or there are psychometric challenges around that, like our primary endpoint, what is the power needed to assess that variable?

There's issues of non-response

bias versus sampling error and there's been challenging -- there's been challenges across the industry representatives on the panel in terms of whether they have been allowed to capture sociodemographic and other contextual data as part of their assessments in order to try to understand the outcome of interest and the variables that may be affecting them.

Furthermore there's some difficulties in surveying before there's high penetration in the market, so the recruitment period may actually be prior to the actual uptake of a product by prescribers.

In terms of recruitment issues, there can be HIPAA considerations if you are using healthcare providers to recruit, and also there are many challenges in terms of industry compensation for healthcare providers doing this work.

And finally, the last two bullets on this slide, refer to the challenges of representative patient populations, which I

think Dr. Shibuya was just mentioning, that there's some initial work that suggests there might be some differences in terms of what we actually capture in surveys than the representative patient populations.

The challenge is however that is sometimes you can have initial drug uptake in which there is a changing array of sociodemographics to those using that product in the early months to the later months, or you can have orphan drugs in which that population is very small and you can only -- you can't assess the same group that you have assessed before, so you start to have limitations in terms of who your sampling frame is.

So that is where I'm going to end and Dr. Smith is going to take over. Thank you.

MODERATOR TOIGO: Thank you Dr. Holmes.

PANEL MEMBER SMITH: So, good

morning. My name is Meredith Smith. I am from Abbott Laboratories. I'd like to start off by, again, thanking the FDA for convening this workshop and also for providing industry the opportunity to share their perspective on using surveys for the purposes of REMS program impact assessment.

As Bill mentioned, these comments, the following slides are really going to touch on a number of additional issues around survey design, administration and interpretation of survey results.

We will have some comments as well about additional sources of data that can be used in this regard.

I think you should be clear, though, that we are offering up these comments in two spirits. One of them is in terms of lessons learned to date, and then the second is in terms of unresolved challenges that we continue to struggle with and that we don't necessarily have a solution

for but like to put forth to bring to the FDA's attention.

One of those first kind of challenges is that involved in surveying special patient populations in particular, whether due to the medication or the disease condition, or to the ageing process, there are certain patient populations where we see varying levels of cognitive impairment.

And other challenges, when the patient population is very small it's difficult not only to identify patients, but for the purposes of the continued cross-sectional assessments over the 18-month, 3- and 7-year period, it's hard to avoid re-surveying many of the same patients.

Another issue has to do with language differences. In some patient populations there may be a sizeable portion of them that are not English speakers, and it may be more than one language other than English, so that's created some other

1 challenges.

2 And then finally, for those

3 patient groups where they are from different

4 cultures, they may be fluent in English, but

5 there are continuing differences in cultural

6 interpretation of certain questions.

7 So given these issues, it's

8 critical that industry has the time to

9 consult with patients, consult with experts,

10 to then pre-test and revise the

11 questionnaire, and that's often a real

12 challenge.

13 It's particularly challenging

14 because we want to get feedback from the FDA.

15 We have a very limited time period in which

16 to do all this pre-work, submit and then wait

17 for industry -- for FDA comment.

18 Some of that comment can be -- are

19 not always timely unfortunately and the

20 feedback can be inconsistent. For those of

21 us in industry who have multiple drugs that

22 are across therapeutic areas that are in the

process of different stages of REMS assessment, we look at the comments and they are just not always in line with each other, so it makes if confusing to understand the FDA's position in regard to certain issues and questionnaire design.

One size does not fit all and we really want to emphasize the importance of being able to be guided by best science in terms of designing a questionnaire as opposed to having to deal with a standard template for assessment purposes.

Just a couple of points about survey administration. Probably one of the biggest challenges to date has been in terms of obtaining a representative sampling frame for patient sample selection purposes.

And this is very, very difficult and as has been noted previously, there's often, as a result one has to resort to some type of convenient sampling.

There are internet patient panels

that are out there for sale, but we know that they are not always representative, and indeed in some cases, they might not even be including the legitimate patients.

One of the learnings we have had that has been useful in some instances is in terms of the actual risk communication materials, using a photograph to anchor respondents.

What we have learned is that both healthcare providers and patients receive risk communication information from a variety of different sources.

And for patients, many of them aren't familiar with the term Medication Guide. They get things from the pharmacy. They may get things in other ways as well, and if they are just not attuned with the concept of what a Medication Guide is, so including a photograph or showing a prompt, a visual prompt of what these documents are, can be helpful in terms of anchoring the

patient and the healthcare provider and jogging their memories.

Another challenge is the fact that, that for -- particularly for the patient Med Guide assessments, an open book approach has not been allowed.

So patients are required to complete the survey without referencing the actual Med Guide itself. And we would argue that this isn't a real world scenario, that Med Guide is not something that patients should be expected to memorize from front to back.

It contains a great deal of information, some of it very, very detailed, and what we would be more interested in seeing is whether or not this is a document a patient can navigate successfully, that they can comprehend the concepts and information presented there, that they retain the document and that they refer to it over time, during the course of their treatment

experience.

So we would argue that not being allowed to use the Med Guide in these types of assessments is not providing us with the information that we'd like to know.

We have also learned that there are a number of biases inherent in each of the possible data collection modalities for survey administration, and to use a variety of them is probably the best we can do in terms of mitigating the different types of biases that are associated with each one particular method.

And then another challenge that has been ongoing is the reluctance of prescribers to participate, so we have a high non-response rate from them, which of introduce significant bias.

So in terms of survey results, probably the biggest question for us is how do you define success? What is that threshold that indicates that our efforts at

educating patients or healthcare providers has been effective?

We would argue then that it's not mastery of specific facts as we often see in some of these questionnaires, but really kind of mastery of the gist, the essence of the information.

We also struggle with the importance of measuring behavioral intent and behavior but with the recognition that self-reported data may not be always the best way to obtain that.

Clearly behavioral intent must be self-reported, there's no other way to get it, but in terms of behavior itself, we should look to other additional supplementary sources of information as well.

As Bill noted earlier in the previous slide, we don't have baseline data, so there's no pre-data to determine if knowledge actually improved, which is very -- makes it extremely difficult then to

interpret whether or not this particular risk communication vehicle has contributed to a patient or healthcare provider's understanding level.

And then we also struggle with interpreting results because of the confounding by variability in terms of the length of the recall period for any given patient when they are recruited into the survey, their degree of exposure to the material -- are they a new patient, have they seen this Med Guide just once or have they received the Med Guide and had an opportunity to look at it over an extended period of time?

There's also great variability in the extent to which respondents actually read the material, whether they read some of it. And then importantly, the extent to which an individual might perceive the information in the Med Guide, all the information, as being personally relevant, and that's important

1 particularly for products that have more than
2 one indication. So patients who have one of
3 those indications may not attend to the
4 information about other uses of the product.
5 I know Med Guides are acknowledged
6 to be patient-friendly but in fact they are
7 not, and there's a substantial body of
8 evidence to suggest that they are not.
9 Incorporation of verbatim language
10 from the Med Guide into the questionnaire can
11 be very difficult for patients to understand
12 and interpret, and we found that again and
13 again.
14 A patient's self-report -- we are
15 also asked as part of these assessments to
16 determine if patients actually received the
17 Medication Guide. That's part of the REMS
18 assessment process, and we know that patient
19 self-report is not always the most reliable.
20 We can think of examples where the Medication
21 Guide is dispensed in the carton with the
22 medication. If the patient is actually

taking the product, they have physically received the Medication Guide. We know that, and we have other ways of tracking that.

But we then look at a patient's self-report for that same product and say that a certain percentage claim they never got it, so we have that issue about relying on patient self-report alone as a measure of that receipt.

And then there's also, in terms of interpreting results, a kind of a paradox that we have seen with some responses from healthcare professionals, which is they score very high in terms of knowledge, but their beliefs and their actual practice behaviors may be at variance with that, and we know that from looking at other data sources.

So just a point, in terms of thinking through whether knowledge is a primary focus -- should be the primary focus in this regard.

Here's a couple of comments and

thoughts about additional data collection options.

These are suggested as supplementary. Drug utilization studies can be a good source of information on physicians' prescribing behavior.

Patient registries can be used to -- particularly if they have a comparator arm build in them, can be developed that can accommodate a variety of different endpoints and provide a richer source of information about the patients.

There's a host of different secondary data sources that also can be used. I'm thinking of electronic medical records, a variety of different healthcare claims databases.

And then we have, increasingly have access to some excellent data from patient web-based communities such as one by Patients Like Me.

Of course there are limitations

with each of these data sources, and some options may or may not be appropriate, depending on the nature of the specific REMS, but we would encourage consideration of these additional sources to give us a broader and a richer, deeper sense of the impact of these risk communication vehicles on providers and healthcare professionals and patients.

So just a few concluding comments. We would really argue that to assess the full impact of risk communication for REMS, is that we go beyond just thinking of knowledge, include additional constructs.

It's important to be able to collect socio-demographic information on respondents so that we can understand differences how well -- how representative a particular sample is in any given REMS assessment instance, but also to understand that there are particular subgroups of patients for which the REMS risk communication has not been effective or has

lesser effectiveness.

Questionnaire design administration should really proceed from the best scientific practice in the area. And that would mean necessarily that standard approach to questionnaire design may not be appropriate, and that there be consideration given to the value of using an open book approach for a Med Guide assessment.

There's a real clear need for public health consensus in terms of what is the primary endpoint we need to be focusing on for these types of surveys, and what the criterion for success is.

A multi-methods strategy for evaluating REMS' impact would be a very valuable addition to the current approach.

And finally we'd like to conclude with a plea for using best design practices for some of these REMS risk communication materials, so that the cognitive demands, particularly on patients, would be decreased,

and these would be consistent with good, plain language principles for risk communication material design.

Thank you. I'd like to ask if any of my fellow panelists have any comments from industry.

PANEL MEMBER STEMHAGEN: Annette Stemhagen. I just wanted to make a comment because one of the other things we're very cognizant of in terms of the questions of timing and all the work that needs to get done to get through that 18 months is of course you need some data on whether the REMS is working because they have the public health of the product in the marketplace.

So we certainly were thinking about that as well as we were developing our comments here.

MODERATOR TOIGO: Thank you. Are there any clarifying questions for Drs. Holmes and Smith, that can't wait until the panel discussion, since they are part of the

panel?

(No response)

MODERATOR TOIGO: Okay. Then we'll now move to the first open public session on -- and both the FDA and the public believe in a transparent process for information gathering and decision-making at FDA, and we place great importance on the open public session part of our meetings and workshops, because the insight and comments provided by speakers at these sessions can help the panel members in the consideration of the questions that we are posing today.

And so, to facilitate a smooth process for this session of the meeting, individual speakers have been allocated three minutes. In instances where multiple people from one organization have signed up, we requested group presentations and in that case gave people a little less than three minutes each.

We'll turn the timer on when you

begin speaking. A light will come on when you have 30 seconds left, and that will signal you to wrap up. I'm not going to use the buzzer and I'm not going to cut you off, but I am going to monitor the time and I'll nicely ask you to finish up if you are going over the time.

And then I will ask the speakers, the panel members, if there are any clarifying questions, again being sensitive to wanting to move the meeting along, but really, if there is a question, we don't want you not to be able to ask one of the public presenters.

And the speakers will be advancing the slides themselves. Is that correct, Cheryl? Okay. So, our first presenter, then, we have three presenters from Covance Market Access: Edgar Adams, Jessica Brainerd, and Shoshana Daniel.

And is that Dr. Adams, the first speaker?

74

DR. ADAMS: Yes.

MODERATOR TOIGO: Okay, thank you. I didn't -- I tried to meet the people before, and I didn't meet you. So welcome.

DR. ADAMS: Thank you. Okay, good morning. I'd like to thank the panel and the FDA for giving us this time to speak. I'm Edgar Adams, Executive Director for Epidemiology at Covance.

With me are Dr. Shana Daniel, a bio-statistician and Senior Analyst and Jessica Brainerd, who is a Risk Operations Manager.

Today we are going to discuss, albeit briefly, the use of multiple choice responses in case studies, the representativeness of samples and knowledge thresholds, and survey assessment timelines.

We do believe that true and false questions in multiple choice are adequate to measure knowledge if properly designed, but we think case studies can be good for getting

at concepts such as the application of knowledge.

And what I have here is the type of question that one might ask. So, for example, your patient is taking eight milligrams of hydromorphone Q 4-6 H. The patient decides that she's taking too much and stops the medication on her own and stops it immediately.

After some period of time, the patient calls you to complain about feeling anxious, restless, having chills, fever and perhaps cramps.

And so the question could be how would you treat this patient? You could use an open-ended response but that would require you to read it or do a keyword search, or you can use multiple choice responses.

So you could counsel the patient, you could have a response of tapering them further from the medication, or reintroducing the medication. But we think that this will

work well.

This type of question, though, needs to be carefully pre-tested with experienced physicians in the area, to make sure you are capturing the entire concept correctly.

As an example, in some cases the symptoms of the onset of the condition mimic for example anaphylactic shock and pre-testing in this type of case would be critically important.

DR. DANIEL: Thank you Dr. Adams. This is Shoshana Daniel speaking. Using a multi-modal recruitment method can allow us to target individuals across socio-economic and demographic spectrum, and also allows for inclusion of individuals who may have media preferences, individuals who may prefer to complete a survey online, by paper or even by phone.

This will allow for a diverse sample. However there's often sparse

available knowledge on the underlying population. Without knowledge of the underlying population, we are not able to extrapolate the results from the sample to the population at large because the sample may not be representative of the population.

The good news is that we have seen time and time again through numerous REMS assessments that such criteria as educational attainment, years in practice in the case of a prescriber or physician, does not appear to have a direct association with knowledge rate.

In fact there is no apparent trend that we have seen, and individuals with lower educational attainment may actually score higher on these surveys than those who have postgraduate degrees.

While these factors do not impact knowledge, we do feel that there are certain characteristics that should be taken into consideration when establishing the threshold

at which we assess knowledge.

For instance, a threshold of 80 percent for a patient population may be appropriate. But prescribers may arguably have a higher threshold, say at 90 percent, as we expect them to be able to better convey this material, the risks that should be conveyed in the REMS, to their patients.

This, in conjunction with other such factors such as the age of the product, the severity of the risk message, should all be taken into consideration when establishing the threshold at which we gain success.

MS. BRAINERD: Thank you Dr. Daniel. I am Jessica Brainerd, and there are numerous factors that impact how long it takes to conduct survey assessment, and when I say conducting a survey assessment and the length of the project duration, I am talking about a soup to nuts approach, including the design of the survey, the survey protocol, the FDA review time which we have heard is at

least 90 days, pre-testing the instrument, conducting the surveys, conducting your data analysis and then writing the REMS assessment reports.

In our experience, on the short end we have conducted this entire assessment over a nine-month period of time, and on the longer end, in rare instances, up to 16 months.

And this all applies to assessments due at 18 months or 3 years following the REMS assessment, unlike some of the shorter assessment periods of 6 months or 12 months.

Building of the surveys and how long that can take certainly impact this duration. We, depending on the various products, have fielded our surveys in as little as 13 days and as long as 8 months for lower-prescribed products and orphan drug products.

Factors that do impact duration or

other factors including both the client and the FDA review time, and needing to factor in at least 90 days for the FDA review team, but our clients need ample time to review drafts of the survey protocols and the assessment reports and that can vary from client to client.

Also, we employ numerous recruitment strategies, depending on sponsor preferences, and sometimes there are reasons at the sponsor level that they do not always want to employ what we would consider the fastest recruitment strategy, so it can also impact the length of time.

Whether a product is an orphan drug product, an orphan drug or a widely-prescribed product, can have a large impact and especially because it will impact the overall population from which you can draw your sample, and it can result in longer recruitment periods.

And finally, whether a product is

new or already marketed, if it's already got deep penetration in the marketplace or if you need to factor in market uptake for a newly-launched product, can have an impact on the duration.

I do want to just touch on respondent remuneration. One, we do believe that this has an impact on the success of the REMS in terms of meeting sample size, but we also do believe that it is of value to offer this to both patients and prescribers, in recognition of the time expended, and these are modest remunerations that are simply commensurate with the level of effort that the patient or the HCP has spent on taking the survey. Thank you very much.

MODERATOR TOIGO: Thank you. Any questions from the panel for our speakers? Any clarifying questions or --

(No response)

MODERATOR TOIGO: Okay. Thank you to the speakers from Covance. Our next

speaker is Dr. Jack Henningfield.

DR. HENNINGFIELD: Good morning. I am Jack Henningfield, Vice President of Research, Healthy Policy and Abuse Liability at Pinney Associates, where I consult on pharmaceutical development and regulation and that includes advising pharmaceutical clients on the development of risk management programs and REMS, post-marketing surveillance and so forth.

I will focus on two challenges that must be addressed to collect data that will be useful in regulation, including communications and REMS evaluation.

The first challenge is to clearly identify the issues of concern, the populations to be assessed and the behaviors of interest.

The second challenge is to apply psychometric scaling science, so that the resulting data are relevant and interpretable.

Defining the issue of concern is vital because there is no one size fits all template for questionnaires. The structure, the format, the reading level of the questionnaires, will depend on the population, and populations include doctors, pharmacists, on-trained healthcare providers in the home, and the patient populations, which can be incredibly variable in education and cultural background.

Yet, we have seen questionnaires designed as though the populations were all the same and people were all linguistic Oxford English Dictionary scholars.

The questions raised by the drugs, and the safety issues also, vary incredibly widely. For example, preventing pregnancy is critical with isotretinoin, whereas secure storage and disposal is critical with fentanyl patches. So the issues, the drugs, are extremely variable.

Assessing knowledge of safe use and

behaviors of interest cannot be approached the same way for each drug and each concern. Once the issues are defined, we need to apply psychometric science to scale development.

This can add cost and time but if it is not done, the resulting data may be uninterpretable or misleading.

Scientific questionnaire design emphasizes the use of minimally ambiguous questions that have been demonstrated to be appropriate to the population, to provide unambiguous response options.

Pre-testing is vital to ensure an acceptable level of reliability and validity, and to develop an estimate of the variability that is expected in the target population.

Finally, FDA can contribute to this process by ensuring that sponsors clearly define the priority, key issues of concern, and verify that the resulting data will be useful in the development and the subsequent evaluation of the REMS.

FDA should question the sponsor to determine if there is acceptable scientific foundation for the approach, and if there is significant confidence based on pre-testing.

This is essentially the role the FDA takes with other evaluation studies for reliability for safety and efficacy, where FDA doesn't design the test, it makes sure the sponsor is doing it right, and it's the same for FDA here.

MODERATOR TOIGO: Thank you Dr. Henningfield. Any questions for Dr. Henningfield from the panel?

(No response)

MODERATOR TOIGO: Thank you for your presentation, and while Cheryl is getting the timer ready for the RTI Solutions, we had two new panel members join us, I think after we did the introductions. Dr. Krotki, could you introduce yourself please?

PANEL MEMBER KROTKI: Karol Krotki,

RTI survey statistician.

MODERATOR TOIGO: Okay, appropriate timing for the introduction here then. And on the end there.

PANEL MEMBER LEFKOWITZ: Doris Lefkowitz from the Agency for Healthcare Research and Policy.

MODERATOR TOIGO: Okay, thank you. And the next presentation, we will have eight minutes for RTI Solutions. We have three speakers: Sandy Lewis; Kelly Hollis; and Laurie Zografos. I hope I got it right.

MS. HOLLIS: Okay, thank you. My name is Kelly Hollis. I'm Global Head of Surveys and Observational Studies at RTI Health Solutions. My colleagues and I appreciate the opportunity to speak with you all today.

A major topic for discussion among researchers involved in REMS surveys is the establishment of knowledge thresholds for gauging whether a REMS educational goal has

been met.

Consideration has to be given to whether an established rate of knowledge may be applied across all REMS programs, or whether thresholds should vary based upon the unique characteristics of a program.

While there's clearly a need to define performance measures, establishments of standard thresholds applied to all REMS assessments may not allow for flexibility to evaluate the impact of educational materials for those who are greatest at risk.

Rather, programs should be considered individually to determine the appropriate threshold for knowledge of key messages, based upon the individual characteristics of a program, such as the clinical importance of the risk the nature of the underlying condition, and the relevance of the risk to the target population.

At this point, a priori thresholds are rarely, if ever, based on the evidence or

the minimum knowledge necessary to minimize product risk, and frequently we find that knowledge may be high among particular subgroups for whom the risk is most relevant.

In a recent survey, we found that the sample population overall had relatively low awareness of a particular risk, but when we explored knowledge among the subgroup for whom the risk was most relevant, knowledge levels were high.

The program overall may have failed to achieve a threshold, if one had been set a priori, when in fact those who were most at risk were knowledgeable about that risk.

Furthermore, knowledge and for that matter exposure to risk, may vary based upon many other factors such as the severity of their disease, severity of potential risk, chronic versus acute illness, physical or cognitive impairments, length of time on treatment, counseling by an HCP, involvement in an ETASU program. These factors should be

captured and analyzed to more fully understand the impact and success of a program's educational initiative to minimize risk.

If thresholds are established, either for a particular program or across the board, consideration should be given to whether those thresholds are established only for knowledge of a primary risk or whether it's equally important to set rates of knowledge of safe use requirements and signs and symptoms of side effects.

Furthermore, assigning a pass/fail criteria for questions in which multiple responses may be correct might lead to an underestimate of knowledge. Care should be given not only in constructing these questions, but also in their analysis.

Finally, it's important to remember that knowledge of key messages is the measure of only one aspect of the REMS program. Minimization of risk may be achieved through,

for example, a physician or other HCP intervention, despite poor knowledge among patients.

Therefore, results of awareness surveys should be evaluated in the context of the overall REMS program.

MS. ZOGRAFOS: Hello, Laurie Zografos, also from the surveys and observational studies group. With sampling for REMS evaluations, one of the many goals is to achieve a sample that is as representative of the target population as possible.

Unless you have access to the entire target population, such as when the drug is part of an ETASU restricted registry, patients will be recruited from a source representing only a subset of all patients taking the drug.

Throughout the process of identifying the sample that is ultimately invited, there are several opportunities for

selection in respondent bias to be introduced.

So some considerations when sampling. First, the sampling source should provide diverse representation and broad coverage. In some cases the best sources for targeting that particular population may be inherently biased, for example patient support groups may provide easy access to patients but the patients may have more severe disease, or have received additional education about their medications.

It's also important to consider the reliability of the eligibility data available. For example consumer panels may rely on patients to self-identify, versus having patient eligibility confirmed by a physician or pharmacy.

Frequently, companies are expected to include multiple sources in an effort to expand representation. However, if the individual sources do not effectively

represent the target population, the combined approach may not be any more representative than a single source.

Respondents could be different from not non-respondents or from the target population in important ways, making an analysis of potential selection or response bias an important step.

When possible, perform a comparison of the distribution of respondents to the expected distribution of the target population and ideally perform a comparison of the characteristics of respondents and non respondents. Of course, the ability to do this comparison and the level of detail available for respondents and non-respondents will depend on access to the individual level data.

In conclusion, it is important to design REMS assessment surveys in a way that minimizes selection respondent bias and allows for some comparison of the

characteristics of respondents to non respondents and/or the target population of the REMS when data are available.

MS. LEWIS: Good morning. I am Sandy Lewis. I am the director in the patient-reported outcomes group at RTI. And I'm going to be talking about the role of cognitive pre-testing in REMS assessment surveys.

Cognitive interviewing is an integral part of a REMS assessment. Because of the importance of these interviews, use of a standard, rigorous approach is essential.

Cognitive interviewing was developed to optimize instructions, question wording, response options, and overall format. The cognitive interview allows us to understand the way that respondents comprehend items, how to retrieve information and select a response.

It allows us to see if respondents are interpreting questions and responses

consistently and aligned with the developer's intent. So, are we getting the information that we think we are getting and need?

Cognitive pre-testing prior to fielding a REMS survey will optimize data quality, contributing to the overall success of the program.

Cognitive interviewing uncovers problems in even the most well-developed questionnaires. Fielding a survey prior to cognitive pre-testing leaves sponsors open to the risk of including poorly-written or confusing questions, leading to challenges when interpreting results.

Examples of important participant feedback from face-to-face interviews include identified language that was perceived as frightening but could have impacted patient behavior.

A discovery of a lack of understanding of key safety information we found was due in part to patients feeling

overwhelmed with the amount of paperwork being received.

And these same patients indicated that the outcome, if they did not take the medication, was much worse to them than the risks that were associated with taking the medication. Thus they were not reading the medication guide that was provided.

Also, implementation of changes to simplify items to facilitate understanding by respondents with cognitive deficits. Additionally, non-verbal cues that can only be observed in a face-to-face setting allowed greater insight and opportunity to probe more deeply to uncover otherwise hidden information.

So in conclusion, a variety of resources describing best practices for cognitive interview methods exist. Ensuring that cognitive pre-testing is conducted by highly-experienced interviewers, who are trained in survey design, questionnaire

development and qualitative research methodologies is essential.

The inherent value in cognitive pre-testing underscores the priority for inclusion of this step in REMS assessment programs.

MODERATOR TOIGO: Thank you to our speakers from RTI. Any questions from the panel?

(No response)

MODERATOR TOIGO: Okay. Our last -- yes, our last presenters are from the Institutes for Behavior Resources, Dr. Steven Hursh and Dr. Peter Roma.

DR. HURSH: Good morning. I'm Steve Hursh. I'm the president of the Institutes for Behavior Resources and what I'd like to talk about this morning is a relatively general approach to how to understand and approach risk management. The Institutes for Behavior Resources has been providing guidance for risk management

strategies for several federal agencies and we follow a process which, like the scientific method itself, is based on precise definitions, valid metrics and continuous cyclical improvement.

These kinds of conceptual frameworks for risk management have been successfully applied by the Federal Aviation Administration and the Federal Railroad Administration and through the application of the approach that we have provided have had dramatic impact on regulatory requirements associated with fatigue risk management in aviation and the rail environment.

The approach is based on a five-M approach which involves measurement of current conditions, modeling of those conditions and projection of potential risks, development of a management system, and it generates mitigations for those risks, and then followed by a monitoring of the outcomes as appropriate to those systems to determine

if the mitigations have been successful.

The challenge for this and it's center to this, is a consensus of all the stakeholders and a model that provides an accurate measure of the risks associated with the current conditions.

Based on this and the challenges that are associated with it is considerable public and private investment in developing the models that drive this system reliable and valid low effort techniques for objectively assessing the current conditions and the appropraite -- and the associated risks, development of models that will aggregate all of the information from the various conditions that lead to risk and then provide unified metrics that can lead to appropriate decision-making.

I'm going to pass it off to my colleague Dr. Roma, who will talk more specifically about an aspect of this whole process which has to do with understanding

behavioral intent. Knowledge is one part of this process. Understanding what drives and motivates behavior and intent is another component, and there are specific techniques that we are working on that address that issue.

DR. ROMA: Thank you Steve. My name is Pete Roma, and we are here to discuss behavioral economics as a conceptual and methodological framework to help inform the risk assessment process. In this case we are sort of focusing on abuse liability assessment and surveillance as an exemplar.

There are basically two schools of behavioral economics. One you might be more familiar with is a sub-field of economics proper. This is where they apply cognitive psychology concepts to understand the irrational financial decisions that humans make, and you might consider this the behavior of economics.

Our approach is a sub-field of

psychology so this is where you apply economics concepts to understand behavior, reward and choice across species, and you might call it the economics of behavior. It's not necessarily limited to financial decisions in the marketplace.

The core concepts of supply, demand, consumption, choice, budget price, all of that apply, except we apply it to all variety of reinforcers, it could be food, drink, drugs, social interaction, anything that's of value, and that is the heart of our approach to behavioral economics, is that money is not the only thing of value. It could be price, it could be time, effort, distance, opportunities lost, anything that is of value to a person can be considered price.

Once of the approaches that we have been -- it's not brand new, but we have been developing new uses for it and new applications of it, is the demand curve, and

101

this is essentially where you plot consumption as a function of price. Price increases, consumption decreases in general, and through that there are a variety of standardized metrics that you can derive that describe -- you can quantify the elasticity of demand, how it changes over price as opposed to choice at any one particular price point, as well as the total output, and this would work for, again, anything that's of value, be it time, effort or money, this could actually have market value itself.

Using the demand curve technology and things like it, there are multiple standardized metrics. You can separate pharmacological and biological effects from the motivational effects. This is essentially a motivational assay. This could be used throughout the risk assessment process, the risk management process, preclinical lab, self-administration to animals as well as humans. There are also

hypothetical demand curves. This can -- this taps into sort of the behavioral intent or at least behavioral possibility, using the same technology and unified metrics across the challenges, of course, that -- the foundations are in place, there is plenty of work that is being done on this. The advantage is with additional support to develop is you can have a quantitative risk-benefit ratio for any -- across classes of drugs, across species.

You can adapt the questionnaires for different populations, for different products, using the same economic concepts of choice relationships, such as substitutes, complements, you can quantify abuse liability in monetary and non-monetary terms, and ultimately you can have this common language of economic concepts and metrics throughout the risk management process.

MODERATOR TOIGO: Thank you Drs. Hursh and Roma. Any questions from the

panel?

(No response)

MODERATOR TOIGO: Okay, well that concludes our open public session, and according to the agenda, we have time for a 15-minute break, and 15 minutes is what we will have so I need you to be back by 9:45 or a little bit before, and for anyone who -- housekeeping detail, for anyone who has luggage in the back -- never mind, we are going to cancel that housekeeping detail, so somehow it's been resolved. Thank you. See you at 9:45. There's coffee and drinks to buy.

(Whereupon, the proceedings in the foregoing matter went off the record at 9:29 a.m. and resumed at 9:48 a.m.)

MODERATOR TOIGO: I have one housekeeping, one more housekeeping before we get started. I had a question during the break as to whether or not we were taking questions from the public for this meeting,

and while I think that would be a good idea, I think you can see from the agenda that we are not going to -- we are going to have trouble getting through the questions that we already have.

At the same time, if there are questions from the public that you think we should have asked in the preparation for this meeting and we didn't, then I think that's important for us to know.

And so I'm going to -- there's two ways we can handle this. We aren't going to really have time, unless things go really, really well, for a public -- for a question and answer from the public.

But what we will do is, you can submit the questions to the docket or if you have them and you want to submit them to me, and I will try at the end either to read all of those questions into the record, or make sure that they get put into the docket but I don't think we are going to have time to

discuss the questions.

At the same time, the person that raised it, what was important is that the question gets out there and that's often very important.

So that's how we are going to handle this part of the workshop. And okay, so if you look at your agenda, now starts the hard part. Our goal is that today's meeting will be a lively and productive discussion about social science methodologies involved in assessing REMS.

We have a large group of experts in the field and we want to ensure that everyone has a chance to participate in today's discussion. Occasionally -- and I'm no expert, so -- but occasionally I am going to jump in on the discussion and maybe ask people to wrap up their comments just to ensure that we keep the meeting moving.

We are also going to try and keep the meeting focused on REMS, on issues

related to REMS, since these questions lend themselves to discussion in a lot of different directions, we are going to try and keep the meeting focused on the REMS.

So I apologize in advance if I intrude sometimes, but I -- if you bring up an interesting issue that we think is not exactly relevant to the REMS discussion we will put it in a parking lot and if we have time we'll get back to that.

And I'll be looking for guidance from the people from FDA who are experts in this area to help me with this part of the agenda.

So here's how we're going to do this. I am going to recognize -- I'm going to read the questions. We are going to go through the questions one by one. I'm going to recognize speakers. If you raise your hand and you want to participate, want to comment on the particular question, Doris Auth next to me is going to keep notes and

make sure that I don't miss anybody.

Sometimes I, at the direction of my colleagues over here, may call on a particular panel member to provide some input if we haven't heard from them and we want to hear what they have to say.

So my goal is to provide some structure for an organized meeting while allowing for some flexibility to help ensure that we do have a productive meeting at the end of the day.

And at the same time, I do need to keep us focused on the agenda and make sure that we cover all of the topics and that's no easy task.

And so one more housekeeping thing that I was signaled from the back, there are people -- apparently we don't have enough seats for the people who want to participate in this meeting, so that's nice to have a party and make sure we have a lot of people show up. But unfortunately the acoustics out

in the hallway are not very good, so we do -- the room we were going to use for luggage, we are now going to use for people. So if there are people who are out in the hall, or want a seat and don't have one you can go to room -- what room is it? Any -- okay. So you can see Georgie in the back and she'll take you to the -- make sure you get a seat. Actually, it's an overflow room and it's room 1404.

Okay, so now let's get started. Well no, actually, are there clarifying questions for our panel presenters, so that's Drs. Holmes and Smith, Dr. Willy and Dr. Shibuya. Are there questions that you had that you wanted to hold, or will they come up during the discussion? Okay. Dr. Lee. Okay.

PANEL MEMBER SUNGHEE LEE: I was wondering if there is any standard requirement or a guidance given to the applicants, drug applicants by FDA for the

surveys that they are conducting?

MODERATOR TOIGO: Is there any standard guidance?

MR. GORDON: Brian Gordon. There is no official guidance. However usually when an applicant is having their REMS approved there will be a standard set of language that includes just general social science principles that we send out to the applicants.

MODERATOR TOIGO: Okay, so I think, Julia, you are going to show the question on the slide. Is that correct? And then I'm going to read the question for the record. As you can see it's a long one.

Topic one is consideration of the primary endpoint. It's actually listed as question number two in the meeting materials, but based on our planning meetings and discussing with panel members, we thought reversing the order for question 1 and 2 would be a more efficient way to do this.

So we are discussing question 2 first. So question 2 says: "Is the knowledge rate, that is the proportion of subjects who demonstrate knowledge of the risk message, the appropriate primary endpoint for a survey?"

Part A is, "What factors need to be considered when establishing the threshold for success for educational elements of the REMS?" And part B is, "Should the threshold for successfully meeting a REMS educational goal be set at a knowledge rate of 80 percent or 90 percent, or should it vary depending on the risk message?

"If it should vary, what should the minimum threshold be for success, and should the threshold reflect whether the product is a new molecular entity or the original biologic product, or an older drug?"

And we have 45 minutes to discuss this topic. So who would like to start the panel discussion on this? Okay. Dr. Willis.

PANEL MEMBER WILLIS: Gordon Willis, National Cancer Institute. Can I ask a clarifying question of FDA staff? To me, this has been a burning question related to what the overall objective is here, your research question concerning knowledge measurement, because it seems to me there are two different designs that might be involved and they are associated with different methodologies.

So far what I've heard, I think, adheres to what I consider to be a surveillance model of knowledge. You do a one-time, cross-sectional survey of, say, patients who have received a drug, and you find out what their knowledge level is, it's 80 percent, 90 percent, whatever.

And so that's challenging enough to do there, to get that level established. But even if you can do that, what you know is, you presume that knowledge is at a certain level, you don't know why. Maybe they know

1 something about the drug potential risk,
2 because their physician told them, their
3 mother told them, an alien implanted the idea
4 in their brain, whatever, or it might be due
5 to the REMS or to your patient education
6 materials.
7 But under that model, to my
8 knowledge there is no way to establish a
9 cause and effect. The alternate model would
10 be what we call an evaluation model where you
11 are explicitly asking the question, was our
12 REMS or our patient education materials, were
13 these what were responsible for the change in
14 knowledge?
15 And for that objective, you are
16 being a lot more ambitious, and this requires
17 a somewhat more sophisticated research
18 design, I believe, where you are now in the
19 realms of something like educational
20 evaluation, where you have an intervention,
21 and you have got to have, as Dr. Holmes, I
22 think, was saying, a pre/post kind of

assessment, or maybe an experimental evaluation with a control group to determine what's the meaning of a certain level that you have measured.

And so my question, just to terminate this, comes down to, when you are talking about endpoints, this has to be understood in the context of what it is you are trying to establish here. Is it enough just to find out that 90 percent of patients know that I might die if I take this medication, or are you explicitly trying to assess a change in knowledge level that can be attributed to this intervention that consists of the REMS program itself, because I think that has important ramifications for any research design here that we might recommend as a panel.

MODERATOR TOIGO: Dr. Willy or Shibuya, do you want to answer that, or comment?

ASSOCIATE DIRECTOR WILLY: So, this

goes back to our overview where we talked about REMS have a goal, and we gave you an example of what a goal is, and that is to inform.

And so that is what the purpose, what we are trying to do with these surveys, is to understand whether the patients are informed. But we acknowledged your second point, and that is that there are many sources for information.

And so at this point in time, we focus on just trying to determine whether they are informed, and gather some information about whether they were exposed at least to our tool, which is the Medication Guide when we are talking about patients, and for prescribers, there can be a lot of other information.

But I think, you know, we could have some additional discussion about you know, whether there are other things that we should consider in terms of you know, the

evaluation models. So I don't know if anybody else wants to share anything, or Brian, your experience?

MR. GORDON: Brian Gordon. Just to get, for, if we wanted to do the evaluation approach as you mention, that we would need baseline data, and when a product is approved with a REMS, there really is no baseline data, and so we would have to maybe -- we would have to use like the first assessment as a baseline or something along those lines. But that's something to be considered in the future.

MODERATOR TOIGO: Dr. Shiffman.

PANEL MEMBER SHIFFMAN: Saul Shiffman. First, let me just comment on the issue we just discussed. I mean it does seem to me that the agency and the public health community is interested in what patients know and not necessarily where they know it from.

It strikes me that if one did otherwise, you would essentially penalize the

sponsor who did a good job of educating providers and patients outside the Med Guide, for example.

Nevertheless, there's another reason to collect that other information, which is that it's process information that helps you perfect your interventions, that is if you discover that certain people don't have the right knowledge because they never got the Med Guide, that means you need to improve its distribution. If people got it but didn't understand it, you need to improve its comprehension and so on.

So that information is very useful for tweaking, perfecting your program. But it seems to me that you don't need to make that causal attribution.

So, to address the question on the table about criteria, it seems to me to be very compelling that the criteria need to be based on the clinical risk, that essentially the question is, what bad things would happen

if a patient or provider didn't have the information, how likely are those bad things and how bad are those bad things.

I want to point out that that's the basis that's laid out in a guidance in another part of the agency, in the non-prescription drugs office, for evaluating consumer understandings of OTC drug labels where there's an explicit guidance that says that the sponsor proposes and then ideally reaches consensus with the agency about what the primary communication objectives are, and out of -- and that's to be based on clinical risk, and this is not independent of the next question about sample size, because once you have identified what the risks are, you can use that to set a minimum threshold of knowledge, and very important to emphasize minimum, not what we would like to have, but what's the minimum that would really be acceptable from a public health perspective.

Once you have that, then you can

set up a statistical test that asks can my data exclude any values lower than that, in other words, can I show to a 95 percent certainty that if we said it can't be less than 75, that it's really not -- we know it's not less than 75, and that then drives the sample size calculation.

So it's important to recognize that the two are very tightly linked. You need to set a standard, which I am arguing should be based on clinical risk, and that then allows you to set up the statistical structure to determine a sample size, a degree of precision that you need, and what Dr. Shibuya described about what sponsors are doing, like assuming a 50 percent rate, is the sort of calculation you do when you have no target. It's sort of the worst case for precision. You can get much better if you A, have a defined target that you are testing against, and B, ideally if you have some sense of what you are actually likely to achieve.

So I think the endpoints should be few. You should really sort the ones that are really, really important. A lot of Med Guides and assessments suffer from trying to do too many things.

So I would argue you need to select a few that are the most, of most concern, and set a threshold that is based on -- a minimum threshold based on the degree of concern, and from that would follow a sample size.

PANEL MEMBER HOLMES: Can I ask Dr. Shiffman a clarification question?

MODERATOR TOIGO: Certainly.

PANEL MEMBER HOLMES: And it first flows back to the question from the FDA. I understand that question to mean knowledge rate -- so let's say you have a multi-item scale, multiple items asking a particular question, and I think that is inferring that 80 or 90 percent of people get all of those items correct. Is that what that 80 to 90 percent means?

So it's the proportion of people who got all of those items correct, is what you mean by the knowledge rate there?

MR. GORDON: Brian Gordon. Yes. But it could be for a single question if it's about one risk --

PANEL MEMBER HOLMES: Correct.

MR. GORDON: But it's 80 or 90 percent got that risk correct.

PANEL MEMBER HOLMES: And so what I was understanding you to be saying, Dr. Shiffman, is you could have a multi-item scale where the average -- so let's say it's 10 items assessing a particular risk, and that the average across your sample would be 70 percent, that they got 7 of those items of 10 or 55 of those items, so we are talking about different sort of metrics here, and so I think when it gets to sample size, that has effects as well, so we should be clear about what the knowledge rate is that we are talking about.

PANEL MEMBER SHIFFMAN: Good point, and I was not necessarily thinking of multi-item scales or scoring them as you're only right if you got them all right, and it seems to me that considerations of principle are the same there, that is you really need to dig down and say how important is it that the patient get this item and this item and this item, and simply saying they have to get them all isn't useful without thinking really hard about what is the consequence of not knowing this and is this therefore coequal with the second and third item?

I don't think -- I guess I'm saying I don't think there's a formula. The bad news is it's going to require some hard thinking and in part in a situation where not everything is known.

MODERATOR TOIGO: Okay. Dr. Gilsenan and then Dr. Sleath and then Dr. Morris, did you have your hand up? No, okay.

PANEL MEMBER GILSENAN: Hi, it's

Alicia Gilsenan from RTI Health Solutions. I just wanted to comment on a couple of points with regard to setting a metric and a standard for doing that. As mentioned previously in some of the talks, it's difficult to set a standard metric. In some cases, when one is set, it may be arbitrarily set to 80 percent for patients or 90 percent for physicians for example. But the basis for setting that 80 percent or 90 percent has to be, you know, you really have to think about that, does that mean that you'd feel comfortable if 80 percent of the people know this and that's going to decrease the risk that you are trying to avoid.

And so again, it is important to the clinical significance. It's also important in subsets of the population. If you set for this risk we need to have 80 percent overall knowledge, it may not be true that the whole population needs to have that. So it's really important to look at subsets,

and I think not always -- we don't always take that into consideration historically when doing these types of surveys and sometimes you learn about that when you do the surveys, as to who really needs to know the information.

MODERATOR TOIGO: Thank you. Dr. Sleath.

PANEL MEMBER SLEATH: I just wanted comment about your question. I don't think that knowledge rate is necessarily the right endpoint. I think application of knowledge like do you intend to prescribe X or Y, do you intend to take X or Y, that if you just assess knowledge a lot of these questions, just by chance they might get them correct, especially the true/false.

And I also wondered about how you handle -- a lot of these had I am not sure or I don't know, which in a knowledge test, I wondered how you analyzed, what do you do with those? Do you count them as missing?

Are you just reporting the -- you know, does your 80 or 90 percent, how does the don't knows include in that, or how are they included in that, was my question.

MODERATOR TOIGO: Brian, do you want to answer that?

MR. GORDON: Brian Gordon. They'll report the percent that answered it correct, the percent that got it wrong, and the percent that answered I don't know, and when we look at that, we try and make a determination of possibly is there a reason that all these people are answering I don't know to this question. Is it because it's a poorly-worded question? Does the information not apply to a certain group that's answering this way?

So we look at it as we just -- we try and evaluate and see if there is a reason for that large group that's answering I don't know.

PANEL MEMBER SLEATH: So then if

you sum -- say you have scales that are multiple items, how do you then sum them, although I don't particularly like yes/no, but I guess I would also like to bring up that how you handle missing data is very important, especially when scales have multiple dimensions, lots of items et cetera, and in the examples we got, I didn't -- I mean I thought they were scary because they weren't, a lot of them, I also want to raise up why aren't Likert scales possibly used more so that then you could do more in-depth analysis of reliability, validity et cetera.

MODERATOR TOIGO: Did you want to comment, Brian, on that, or no?

MR. GORDON: Well, for the -- Brian Gordon. For the missing data that is an interesting point that we will have to take it back and further discuss our policy on how we handle the missing data.

PANEL MEMBER SLEATH: And I just want to point out that I think missing data

is different than response rate. Those are two very different things. But if you only get a 30 percent response rate, and then half your data is missing for some of your questions, then you are down to a 15 percent usable data.

MODERATOR TOIGO: Okay thank you, and now let me just -- it's Dr. Raghunathan and then Dr. Beatty and Dr. Chowdhury and then Dr. Lee. Okay.

PANEL MEMBER RAGHUNATHAN: Trivellore Raghunathan from the University of Michigan. Now, before we talk about this knowledge rate, I think we also have to discuss about there are different populations that you are talking about. It's not only the patient population, you are talking about healthcare providers population that you are talking about.

And the knowledge also, is it measuring the knowledge with respect to the material that is being provided, or is it

just absolute knowledge that I think they could have gained from all other places?

I think that that needs to be addressed first. If you are really looking at the absolute knowledge then it is irrelevant or you may assume that this material that you are sending is not the only source by which they are gaining the knowledge.

But I think the purpose is, as I understand, is to know whether the material that you are sending them is giving them the knowledge that they need.

If that is the case, then all this knowledge has to be measured with respect to the material that you are sending them. So this knowledge rate discussion doesn't bring in the question about is this with respect to the material that you are sending them or with respect to general knowledge?

MODERATOR TOIGO: Dr. Willy, did you want to comment on that?

ASSOCIATE DIRECTOR WILLY: I think I am, as I stated before, it's difficult to be able to separate out where they are receiving the knowledge. And so when we do our surveys, we are assessing the knowledge that the patient or provider has, and we do try to get information about where they have received their information from. We give them in some cases a list of different materials if it's a healthcare provider.

But we know there's many ways they can get information. So I guess we could discuss that in more detail in terms of how you might do that kind of evaluation, knowing that there can be other sources, but currently, we attempt just to see whether they have the information we think it's important for them to have.

MODERATOR TOIGO: Dr. Ellenberg, you had a specific comment on Dr. Willy's comment before we go to Dr. Beatty.

PANEL MEMBER ELLENBERG: Yes.

Thank you for recognizing me. Is it on the table that you can do the survey on people who you know to not have received the message? You will be sending out this message to a specified group of people. Could you do the survey without having the message sent out? Could you identify those people?

ASSOCIATE DIRECTOR WILLY: I think it would be interesting to hear from other people on the panel as to whether they think that that is a possibility.

PANEL MEMBER STEMHAGEN: Well, in theory the sponsor is supposed to make sure that the Med Guide is given out to every patient at every dispensing, except for certain circumstances.

So we have always used the assumption when we are looking at that, that all the patients at least had access to it. And then there are questions that the FDA has asked to be put in surveys that Dr. Willy

just mentioned about did you actually get the Med Guide, did you read it, did you understand it, did you ask anybody to interpret it for you.

So, I'm not -- it would be very difficult to find people who, at least in theory, didn't have it, unless we'd started that before the Med Guide was widely distributed, meaning do that analysis in almost this baseline or pre-test kind of circumstance.

MODERATOR TOIGO: Okay, I just want to -- before we leave this topic, Dr. Beatty was that your -- did you have a new thing to raise or was yours related to this? A new thing. All right. Let me take any more comments on this. Dr. Raghunathan.

PANEL MEMBER RAGHUNATHAN: Yes, in that case I think, or one way of doing that would be to ask several questions: did you receive the material; did you read the material; and then assess the knowledge.

Then we can create the subgroups based on various segments of the population and then what their knowledge rate is, and that way indirectly then we can measure whether the people who got the material have more knowledge versus the people who did not get the material.

MODERATOR TOIGO: Dr. Morris.

PANEL MEMBER MORRIS: Yes, well you are talking about experimental design, and the answer to your question is there's two ways you can get people who don't -- who aren't exposed. One is to do it before the Med Guide or the REMS is approved, so you can actually do an experimental test, and the other is do it with non-patients, people who you know don't get the drug or don't get the Med Guides so you have a comparison group. Those would be the two ways.

MODERATOR TOIGO: One last comment on this? Did -- Dr. Gilsenan did you have a copy, I mean a comment? Dr. Gilsenan, I'm

sorry, Dr. Gilsenan did you have a comment?

PANEL MEMBER GILSENAN: I was going to say, these questions are specific to the drug at the time of approval, so it's really not possible to do a pre-test about the risk of a specific drug before the drug is available.

You could ask in general about the disease but I think it would be really difficult to get baseline information on a drug before it's marketed.

PANEL MEMBER MORRIS: Why not? I don't understand. You can pre-test, comprehension test for OTC drugs before they're approved. Why can't you do a pre-test before a drug is approved? I don't understand. Why not?

PANEL MEMBER GILSENAN: Well, for example, one of the -- I'm interested to hear what the other folks say -- but one of the examples of the questions of knowledge that you might be asked is are you aware that

taking this medication can increase death, or are you aware that taking this medication may cause cancer?

PANEL MEMBER MORRIS: Yes, so? I mean why can't you ask that question, if they are reading it in the Med Guide. These aren't people taking the drugs. These are people just testing whether they can understand --

PANEL MEMBER GILSENAN: Okay, sorry. I didn't -- I misunderstood your -- the pre-test and certainly yes, you can pre-test the -- whether the materials are communicating the information ahead of time and I think that is actually done in some cases.

MODERATOR TOIGO: I think we are going to move to Dr. Beatty unless this is really important Dr. Holmes, for the -- to close this part of it out?

PANEL MEMBER HOLMES: Well I'd just say that we do measure whether people get Med

Guides or not, and, or Dear Healthcare Provider letter et cetera, but we are not powered to do those sub-analyses and not only would that not answer the question, but the amount of information people get, these would need to be multi-variable regression models to see what all the different things contribute that knowledge and how much the Med Guide actually contributed, in addition to all those other things. So we are talking about big studies with lots of people to do that sort of level of multi-variate regression, and I don't think FDA wants us to do that, and we are not doing those numbers of subjects in studies. So -- just to give that little bit of experiential background.

PANEL MEMBER SHIFFMAN: One last point on this issue, if I -- may I?

MODERATOR TOIGO: Okay, Dr. Shiffman, but --

PANEL MEMBER SHIFFMAN: I'm going to make it really quick.

135

MODERATOR TOIGO: Okay, and that's good because otherwise Dr. Beatty is probably going to come to the end of the table and make sure that I see him and the others who are waiting. So, okay.

PANEL MEMBER SHIFFMAN: Just really quickly. There is actually another reason to do a test in people who haven't gotten the Med Guide, and shouldn't be expected to know anything, which is to understand what the level of either random responding is, so for example with the true/false roughly 50 percent, but also we have seen questions where the answer is obvious on the face of it, and doesn't really require knowledge, and it would be useful to know that.

So it's not a true baseline. It's not a pre-test. It's a way to gauge what the performance is like of someone who actually doesn't know and shouldn't know the content.

MODERATOR TOIGO: Thank you, okay. Dr. Beatty.

PANEL MEMBER BEATTY: Paul Beatty. I guess I'm stuck on a more fundamental question, and I think some people have touched on this in some of the other presentations but I guess I'd like to hear a little more explicitly the way FDA and others think about this, and it comes down to, when we talk about measuring knowledge, I mean, a knowledge rate sounds like a very nice and precise thing, but what kind of knowledge is it that we are measuring, because I can think of at least several ways to conceptualize it.

One would be sort of your instantaneous knowledge, like you have been exposed to some message. Do you instantly know what it means? Can you regurgitate the gist of it? Did you comprehend it?

There are other kinds of knowledge too, though. You might instantly understand it but not have it for any length of time that it becomes usable to you. So are we talking about more retention or durable

knowledge, and if so, how long are you expected to retain the information that you have been exposed to and to be able to use it?

Or are we talking about something more even pragmatic than that? I mean, is it reasonable to think that someone would receive this message and have it, have access to it in their mind, pretty much indefinitely, or is there other ways that they would use it?

For example, having been exposed to it, would you know where to find it again if you had a question about this? Like, now I'm thinking about this medication, I need to find out what the risks are, I've been exposed to this, but where is that information? Because that might be a more realistic way to look at what people actually know.

Now, these are not really trivial questions to me, I don't think, because

really, it gets at what we are trying to do in the first place, but it also gets at the type of questions that you would ask. You might rate very different instrumentation to get at whether you instantly got a message and it was clear. That, as I said folks, is more a matter of comprehension.

Retention might be a different thing altogether, both in terms of the question and the overall approach that we would do to get this.

So I guess I'd like some clarity in terms of what we are really trying to measure here, because a rate of 80 percent for one would mean something totally different than a rate of 80 percent in another, I would think.

MODERATOR TOIGO: Okay, I'm going to turn to my FDA colleagues to comment on that.

MR. GORDON: Brian Gordon. An example would be there's usually some sort of time frame that says in the methodologies

there will be -- for the inclusion criteria, the patient has received it in the past three months, this drug, and we want to ask them do you know, the drug that you are taking, do you know the risks, maybe the symptoms or the action that they should take?

So that's the type of knowledge we are trying to be able to report on. Did that help clarify your question?

PANEL MEMBER BEATTY: Not really. Is that -- is that a constant then? It's always a three-month period or --

MR. GORDON: Brian Gordon. No, that was just an example. The sponsors established their inclusion or exclusion criteria but you have to, I guess, have a big enough window that people -- they'll have enough people to recruit and balance it versus retention.

ASSOCIATE DIRECTOR WILLY: This is Mary Willy. I would just add, I think what I heard you asking is this -- are we looking

for knowledge that is only necessary for a very short time or for an extended period?

And in most cases it's not, again, as you have heard, these aren't all the same, but that we are interested in having them know the risks and most people are taking these drugs for an extended period of time, so this is information they should be able to retain, at least as long as they are on the drug.

MODERATOR TOIGO: Did you want to comment further, Dr. Beatty, or --

PANEL MEMBER BEATTY: Paul Beatty. I guess, I guess it's just -- so there's a variety, I guess. I guess I just don't understand how the guidance is given and how consistently this is being implemented.

So, if in one situation it's -- we demonstrated understanding instantaneously, that basically they read the message and they got it, then that just strikes me as something different than assessing whether

someone longer term is still remembering information or what the gap is between the time of exposure and the time where they'd be needed to recall this information, so --

MODERATOR TOIGO: So I think there is the REMS, the goal, the REMS have a goal. It depends on what the drug is. There's lots of them. We have -- the experience that we are gaining is, you know, it's still a relatively new program, so I think that there isn't a single answer, if that's what I'm hearing from the panel.

And I think maybe we need to move on, or was it -- Nancy, did you have a comment specifically on what Dr. Beatty had raised?

PANEL MEMBER OSTROVE: It kind of builds on that. It's not specific to that.

MODERATOR TOIGO: Okay, well if you were next in line and you were going to make your comment, would you think that it was a new topic or --

PANEL MEMBER OSTROVE: I think it's related.

MODERATOR TOIGO: Okay, well then go ahead and we'll see.

PANEL MEMBER OSTROVE: And then you can just take me off the hook if it isn't. Nancy Ostrove. The whole issue of kind of different types of knowledge, I think, for me it raised it, especially with respect to physicians. I know we have been talking a lot about patients and Medication Guides. But in some ways, physicians are even more important because they are the ones who make the final decision in partnership.

And there's different types of knowledge there, and the way that that's measured, and maybe it's more appropriate to talk about this in terms of questionnaire design, but the way that's measured is important, because okay, a physician can know, this is what -- I know that this is what FDA believes, and this is what FDA is

telling me, and I know that this is what's in the literature, and then there's the I know what is right for my patient under the particular circumstances.

So there's kind of the, well, what should I do, but what should I do then breaks down to well what should I do according to the labeling in FDA, or what should I do according to the patient and the circumstances.

And those are, you know, important distinctions from my perspective that are very difficult to get at, and I think it's important, you know, that FDA think about, well, what -- are we trying to measure what we are telling the physicians, we want them to parrot back what we are telling them, or are we trying to measure something else? Because frankly, they may not even believe us, because there are plenty of them out there -- I see some shaking heads there -- who, you know, who say, well, this is what

the labeling says and this is what FDA says and the manufacturers, well they are only concerned with liability, and the FDA is concerned with what kind of bad publicity they are going to get if there are too many of these risks showing up.

So, that issue, I think, is important to acknowledge as well, to figuring out what kind of knowledge we are measuring.

MODERATOR TOIGO: Thank you. Two good points on that. We were getting focused on Med Guides and this is bigger than Med Guides, and the second is that yes, we do hear often that people answer the surveys because that's what they believe rather than that's what maybe the information says.

So you raise two good points that's good for the discussion. Dr. Raghunathan and Dr. Smith, relevant to this or do we move on to Dr. Chowdhury? Okay, then, please go ahead.

PANEL MEMBER RAGHUNATHAN: Yes, in

reference to what Paul was raising is that I think is also an issue about the saliency, so if I am -- if a person is diabetic, then the drugs are contraindicated for that diabetic, then I think that may be more knowledge that is stuck in my mind rather than the conditions for which I don't have that -- if I don't have that condition.

So I think the knowledge has to be measured in the context of the patient as well so I can't see how one could come up with the one measure of knowledge rate and say that well, they passed the test, or -- because it will be, they'll have passed the test that is relevant to their own situation so I don't see that -- how one could avoid some sort of regression modeling or what, some sort of a segmentation of the population to really measure what that knowledge -- whether they have the right knowledge to understand the risk and benefits depending on their circumstances.

MODERATOR TOIGO: Dr. Smith.

PANEL MEMBER SMITH: Yes, just to follow up on Paul's comment. So I think that what's really critical for viewing this as a public health intervention, this risk communication approach, is the functional understanding, how patients then take this information and apply it in the real world, real world context, how it guides their response and their behavior.

So I think that their functional understanding is probably the most critical piece of all this, is how they internalize that information and act on it.

MODERATOR TOIGO: Thank you. Dr. Hornbuckle and then Dr. Chowdhury.

PANEL MEMBER HORNBUCKLE: To Dr. Beatty's comment, I would just like to reiterate and mention to the panel that the current practice in terms of primary outcomes is to use the Med Guide, and specifically focus on what information is most important

147

for the patient to know.

So our primary endpoint is that we are looking at the Med Guide and that section of the Med Guide. What is important for the patient to know from a serious, safety risk perspective?

And when you do that, and experiences that we have seen, is that if the risk message is a very simple, understandable, very concise risk message around a serious adverse event, and the broader population needs to know it, we have seen that thresholds of 80 percent or more can be acceptable.

But there's a challenge, though. When that risk message is very complex, when the risk message is very complex and you start asking patients about signs and symptoms, you also start asking patients about subgroups, particular subgroups being at risk, we generally see that that threshold varies, because what happens, the broader

population generally would not know about those sub-populations, particularly if that adverse event is not relevant to that patient.

So, also we need to take into consideration the variability of number one, what is the risk message, what is the goals of the REMS, and the relevancy of the sub-population. This has been a common theme that other panelists have mentioned as well.

MODERATOR TOIGO: Thank you. Okay, Dr. Chowdhury.

PANEL MEMBER CHOWDHURY: How to measure knowledge and/or what knowledge do you want to measure or whether we need to measure before or after, whether we need a baseline or not, these are important issues.

But once we decide how to measure then knowing or determining the exact level of endpoint is different for different purposes. For instance, for the overall REMS, it is important from the subject matter point

of view to determine exactly what level of endpoint, whether it's exactly 80 percent or 82 percent.

This has to come from a different level. But for measuring these, for designing a study, or designing a survey, we do not need to worry about the exact level as long as we know some rough level, what endpoint do we need.

For sample size, it will not make much difference whether the endpoint is 75 percent or you know, 85 percent, it's really in survey, we have multiple things to measure and is really sample size is determined based on the most conservative one.

So I think for designing the study, it's not that important to know the exact level of endpoint.

MODERATOR TOIGO: Thank you. Dr. Lee.

PANEL MEMBER MATTHEW LEE: Well, I think some of my points were summarized by

previous speakers, but I just want to point out, when we are talking about knowledge rate, the situation that we are currently presented with, it's almost a testing situation, because the respondents, the patients, the prescribers et cetera, aren't supposed, quote unquote, "have materials in front of them from which to answer the questions."

Is that really a real world scenario, which was one of the topics we brought up in our presentation. You know, is the most important thing that somebody can memorize the risks associated with a product when there's three, four or five risks associated on that Med Guide, or is the most important thing, hey, I know to pull out this Med Guide and I know where to find the risks on this, or if I am not feeling well, I need to go back to that Med Guide and check, is this one of the things I'm supposed to be looking for, you know, or supposed to be

noticing as one of the risks associated with this medication?

And with that, you know, I want to acknowledge some of the other points some of the other speakers made, about there are multiple sources of information out there for patients. You know, the manufacturer, the sponsor is putting out the Med Guide and in some instances additional material such as healthcare practitioner guides, patient counseling guides et cetera, but there's many other sources of information out there -- the internet, pharmacy handouts, pharmacy leaflets et cetera, that can oftentimes provide misinformation that -- are we seeing influence of misinformation going out to the public, you know, influencing responses to these surveys, whereas, let me present a different scenario, if we are giving the Medication Guide to the survey respondent as a pop-up, can you use this piece of information to answer these questions, then

we have a more reliable source of knowing where their information is coming from because we are asking them to use the actual documents we are providing and not just something they may have heard from a friend, from a doctor, from a relative, et cetera.

MODERATOR TOIGO: Dr. Willy I think you were hoping to get some discussion around this topic, correct? And so you have -- Dr. Lee has raised it and we were hoping to get some input from the panel on this particular issue.

ASSOCIATE DIRECTOR WILLY: You're reading my mind. This is Mary Willy. I would like to hear, get some feedback about this, what you just proposed. Rather, you should be allowing folks to have the information in front of them when you are trying to get an understanding of their knowledge.

MODERATOR TOIGO: Okay, we're going to start with Dr. Smyth.

PANEL MEMBER SMYTH: I think it depends on what you are trying to measure. If you are trying to measure their knowledge, then probably not. If you are trying to measure the usability of the Med Guide, then probably yes. So you are measuring two very different things, depending on whether you have that in front of them.

MODERATOR TOIGO: Okay, Dr. Sleath. We are going to go right down here for the people who have comments.

PANEL MEMBER SLEATH: I wanted -- two speakers brought that up, is can you find the information, you know, later, and several of the Med Guides you gave us as examples state it only has to remain on the company's website for three years, and so I wanted to raise that, is why only three years when drugs are prescribed for very long periods of time?

And the other thing is, I think that if you are doing pre-testing, like other

1 people have talked about, that maybe try it
2 both ways, where they have the information in
3 front of them, and then maybe different
4 groups where they don't have the information
5 in front of them. But I do think there is
6 something to can you find it if you need it,
7 and after looking at these Med Guides, I
8 wondered, how easy would it be able to find
9 it on the websites as a consumer?
10 And one last quick comment, is it
11 tends to be dominated, these REMS, by
12 physician groups as to who you have to give
13 this information to, and being from a
14 pharmacy school, I think pharmacists are
15 important because people often have ready
16 access to them, and also consumer groups, you
17 know in cancer or asthma, et cetera, why
18 aren't companies required to deliver this
19 information to them as well?
20 MODERATOR TOIGO: We'll save the
21 comments from Rob or Mary, but I want to go
22 down the -- okay, Dr. Shiffman.

PANEL MEMBER SHIFFMAN: Saul Shiffman. First with regard to the testing, where you actually say to the person here's the material, read it, I'm going to ask you some questions, there actually are standards for that within FDA, again in the non-prescription drugs area for label comprehension.

Now, those are very important but they ask a different question, which is, you know, I mean, make an analogy to drug development. When you develop a drug you not only have the active ingredient, but you test its bioavailability. This is its infoavailability. Can we actually get it from the paper to the patient?

But that's a slightly different question than, I think that's actually very important because it's necessary even if not sufficient, I mean, if the patient can't extract that information, nothing else is going to work, and we have done a study,

doing a label comprehension study of a Med Guide, and I have to tell you the results were very alarming.

So, for some bits of information that were literally life and death, comprehension was 20 percent. So Med Guides need a lot of improvement. We are focusing here on measurement but we also need to attend to how the materials are developed.

But I -- it seems to me that the FDA and the public health community has here a very practical concern, which is, do patients know or can they get to -- do they have the information or access to do the right thing to maintain safety and it's not a question of does it come from the Med Guide, did it come from a TV program, did it come from the sponsor's website.

And so in fact, if there is misinformation from sources, that's a problem. That's not something that you can excuse because our Med Guide was correct. It

seems to me the public health interest is in knowing what the patient understands, and I would include in understanding, knowing where to find it, knowing that it's in the Med Guide, knowing enough to know I should call my doctor or my pharmacist about this because this symptom needs attention.

So, from a very practical perspective, it seems to me we shouldn't care where the information comes from, and we should care about the information the patient has, even if it means that they have to go get it. They don't have to memorize the Med Guide. Some of these patients are on four or five drugs. Keeping track of which symptom belongs with which Med Guide, which medication, is very difficult, but from a pragmatic perspective, what we care about is that they can recognize something as alarming and either know where to get information, or have the information.

MODERATOR TOIGO: Okay, I just want

to point out we are running over but I think this was information that you wanted to hear, Dr. Shibuya, correct?

So I'm going to ask people to keep the comments brief but let's -- and keep your hand -- we'll go around and if you want to talk on this, please be brief but we want to give everybody a chance to comment. Okay, Dr. Raghunathan.

PANEL MEMBER RAGHUNATHAN: Yes, again this closed book versus open book exams analogy is very different in the sense that they are trying to measure completely two different things.

But if you want to really evaluate the -- how people are able to retain the information that is provided in whatever the Medical Guide or whatever information you are providing, you've got to make sure that the material that you are presenting is like at the fifth grade level, and that's the guidance I think we use in informed consent,

developing the informed consent, is that the fifth grade reading level, is enough to really, to understand those informed consents. So that's a guidance that we are to use.

MODERATOR TOIGO: Okay, before Dr. Ostrove comments, Brian, did you want to make a comment?

MR. GORDON: Brian Gordon. One point of clarification. We don't -- we recommend to the sponsor companies that they do not send the Medication Guide or educational materials with the survey to the patients or prescribers, but we do not say that a patient, if they had a Med Guide at their house, they could use it while they took the survey. We just didn't want education materials to come with the survey.

MODERATOR TOIGO: Okay, Dr. Ostrove.

PANEL MEMBER OSTROVE: Being brief is always a problem. It raises different

issues. I agree with everyone on that. But I think that the open book -- if you do it as an open book, the problem is all you're doing is telling -- is finding out whether people can find it when you give it to them. That's something that obviously needs to be done.

But if you do that, and I actually have some concerns about using that as the basis, it raises the issue of at least measuring what the gist knowledge is of people, because they need to understand that there are risks associated with this drug and I need to check the information more often, especially if I have, say, a patient who I seem to remember is at risk with this particular product.

So having, you know, doing the actual knowledge test, with the open book, I think is actually problematic. I think that, you know, now if we could be sure that people had access to it at the time that they needed it, which is one of the things I know the

agency is working on, then what you could do is measure the gist knowledge, just the meaning and you wouldn't have to have all the specifics, because they shouldn't have to have all the specifics. They should be able to have the referenced text that they can get to when they need to, but they need to know when to go to the reference and then, so it has different implications for how you ask the question, what questions you ask.

MODERATOR TOIGO: Dr. Morris.

PANEL MEMBER MORRIS: Yes, Lou Morris. More of the same. I never understood why knowledge rate was the goal of a REMS. I always thought that for the most part, behavioral outcomes made more sense.

I also could not understand why you would wait a year and a half to measure whether people understood the document that you are dealing with. Doesn't it make more sense to do a comprehension test initially, right around -- right before or at the time

of initial approval, as they do in OTC drugs, because then you are -- the value of that isn't so much the test, but the process of developing the Med Guide is the -- all the qualitative work goes -- that goes into it, to make the Med Guide or whatever the document is, a better document.

You really want to maximize your ability to communicate at the time of initial dissemination so why you are waiting a year and a half to do this testing, is beyond me.

So why not do that? And then downstream, if you want to evaluate, evaluate on a behavioral outcome, if you want to measure knowledge that's fine, because that's instrumental and that will help you tease out what is actually going on there.

So I think the goal of the REMS should be much more behavioral and knowledge should be done right up front with a comprehension test.

MODERATOR TOIGO: Thank you Dr.

Morris. This side of the table. Okay. Dr. Cantor.

PANEL MEMBER CANTOR: Right. Actually, Dr. Morris sort of stole most of what I was going to say. I was -- there are really two different things going on. One is do they understand -- can they find the material, and what is their ultimate knowledge once everything has been implemented.

And it seems to me that both for our process reason but also as a test of the material itself there should be some requirement that in the development process there is some set of sessions going on where the people are brought in, they are asked to find material, and get tested on the material, and there has to be one standard for the material itself and it improves the way the material is developed.

That's very separate from how effective the material is once it's deployed.

So when people are asked to report on things after they have gotten the drug as part of the survey, as a practical matter there is no way to control whether they are looking at it so you are really just interested in what the endpoint is, and also as a practical matter, I doubt most people would be able to find it when they do the survey.

So that's just whether there's knowledge or not. Now whether knowledge is an important aspect of this, I'm not sure, but it seems to me that part of the requirement should be doing this initial development work and actually reporting on the results of that development work before the thing is deployed.

MODERATOR TOIGO: Thank you. Dr. Holmes.

PANEL MEMBER HOLMES: William Holmes. If open versus closed book is a key important factor that you want to be exploring, the practical way these surveys

are often done, there may be a mixture of open and closed book, correct?

So if you are doing an internet-based questionnaire completion and you are a healthcare provider, you can have your prescribing information right next to you or online, and we never -- I don't think the ones that I've seen ask whether they have been referring to the prescribing information.

So that may be, not to proscribe what we should be doing, because one size doesn't fit all, but if open versus closed book is really important, we may already be doing predominantly open book testing. We don't know, because I don't think we have been asking that question.

MODERATOR TOIGO: Thank you. Dr. Gupta.

PANEL MEMBER GUPTA: Yes, I just wanted to reiterate, as far as the materials go, I mean I think somebody had mentioned

fifth grade readability. I think that's a very important point so obviously the materials that need to be -- if you're going to test the knowledge, really have to be, you know, the readability and ambiguity has to be tested first so that is one aspect of it.

On the other hand, as far as the survey goes, the questionnaire, the items in there, the construct and content validity really needs to be tested because that is how you really evaluate the information you are getting, whether it makes sense or not.

So there are two aspects to it, you know, readability ambiguity on one hand, especially for the patients, and subsequently, the construct and content validity of the questionnaire that's being administered.

MODERATOR TOIGO: Thank you. Dr. Hornbuckle, Krotki, and then we are going to end this -- I am going to go Dr. Shibuya to summarize so we can move to the next panel,

next discussion.

PANEL MEMBER HORNBUCKLE: Ken Hornbuckle. The current goal of the REMS Is to look at or assess -- is to assess knowledge. To do an open book, as we said, other panelists have mentioned, is more related to comprehension, which is really the outer scope of REMS.

I bring that point up, as others had mentioned, is that there are great opportunities to move upstream and to look at evaluating the cognitive development as well as the understanding, pre-testing of the Med Guide. So, moving upstream and helping the patients understand the information in the Med Guide prior to doing the testing or the surveys, is a very important concept for the panel to consider.

MODERATOR TOIGO: Thank you. Dr. Krotki.

PANEL MEMBER KROTKI: I want to extend this discussion and make a larger

point. So I was getting -- I was having trouble following this discussion and then I realized, that the term knowledge rate sounds much more complex and almost sophisticated than what I think we are really talking about.

We are starting off, the basic -- if I understand correctly, the basic and the simplest scenario would be are you aware that drug X which you are now taking leads to the following risks? Yes/no. And then you'd establish subjective thresholds.

But that's like the tip of the iceberg. I think the whole discussion is about what's under the water. That one question can -- first of all it can, as we have seen from many presentations this morning, can be subject to all kinds of errors.

So even if we go get a value for it, we are not sure exactly what it represents. Heaven knows we certainly

wouldn't call it knowledge.

And then the other part of it is, what are the underlying factors explaining those yes/no answers? I can't imagine that anybody would just be happy with an X percent answered yes, and that would be the end of the story and nobody would want to know about which proportions of the -- which segments of the population answered how.

So I think there is a lot here that FDA needs to decide, as to whether they are going to be happy just living with that one simple question subject to error and no explanation, or how deep do they want to go into that iceberg?

MODERATOR TOIGO: Thank you. Dr. LaVange, did you have a comment that you wanted to add? Okay. Yes.

PANEL MEMBER LAVANGE: Okay. Thanks. I just had several things, back to this question, not specifically to the open book question.

MODERATOR TOIGO: Okay.

PANEL MEMBER LAVANGE: First of all, there's been several parallels to the over the counter and I think there are some parallels there because you typically have a label comprehension. You might also have an actual use study which gets at the behavior.

But it's a little bit different scenario because these are prescribed drugs. With over the counter, you know, the world can buy those.

So we have the advantage of knowing who has the prescription, so we actually have a finite population which we don't always have in other scenarios.

I think that -- but clearly the pre-testing of the instrument needs to happen, and there's all kinds of cognitive testing and other mechanisms that can do that.

But you would be doing that in people that probably hadn't had the drug

prescribed for safety reasons, and so I think the person who has the disease, who needs to take the drug, has got a little bit different attitude they bring to the table when you are trying to assess their knowledge, and so you are limited but you can certainly do some cognitive testing on readability and understanding and so forth.

But just in a simple way, I see this as first of all, there are lots of endpoints but we just were asking about the primary, because that is going to lead to the sample size and some other things.

You know, you have a target population, these people have the drug prescribed, there's the characteristic -- are they aware of the risk, which we are calling knowledge I think.

But it's really more of an awareness and it really doesn't matter in a public health sense where it came from. It could come from a TV ad or it comes from a

REMS.

But in a public health impact perspective, I want to know if the people that something bad can happen to, know that it can happen to them and what they are supposed to do.

And that's just simply a characteristic of the finite population and you sample the population and you estimate it and it has a precision and that that will drive the sample size and other aspects of the study itself.

You know, why they know it, whether it came from the REMS, those are good things to know. It's a more formal assessment of the REMS. But just from a public health perspective I do think this is the right primary endpoint, and we can measure it and we can measure a lot of other things and we can measure it in subgroups that are at higher risk than other subgroups or subgroups that admit they have read the instrument or

not.

A couple of other points, though, that, I think it's important that we do need to set a threshold. We can talk in the next question about whether we are going to just estimate the variance or we are going to try and exclude the threshold.

But that hopefully is done in advance and you may have a different threshold for different subgroups and they may need to be out there in advance and nobody has mentioned that but I think that's important.

And then finally the thing that had me the most worried is that we are measuring awareness, there are people that are not responding and there are people that have not read the package or there are people that don't respond to 3 out of the 10 items on the scale or whatever, and I don't think that's random missing-ness, so to speak. I worry that the people that are not responding are

the people that don't understand.

And so the non-response bias could be big, and how we impute for that and there are certainly ways to do that, is tied into the appropriateness of this primary endpoint so I don't think we can talk about the primary endpoint without also thinking about how we are going to estimate it and what we are going to do with this non-response.

So, and then one last plug is I just want to assume that we are dealing with multi-cultural, multi-language concerns here as well, because there's lots of different cultural beliefs, so you know, knowledge and awareness and different cultures has also got to be wrapped in there.

MODERATOR TOIGO: Thank you. So you said sample size enough times that that means we can start moving on to the next question. I know we didn't get to everybody, but comments to the docket if you really have things you want to say that we didn't get to,

but we are running over and we had allocated more time for this next topic.

So topic number two is strategies for recruitment and sample size considerations, and again, it's actually listed as question number one in your meeting materials.

And I'm just going to read it for the record: "What strategies can the applicant use to recruit a sample that is representative of the population that is prescribing/dispensing/taking the drug?

"Part A: given that the applicant cannot compel an individual to complete a survey, is it acceptable to enroll a relatively small, making the survey feasible number of participants that are representative of the totality of the healthcare provider or patient population and make generalizations from that sample to the larger population?"

And part B of the question: "What

is an adequate sample size to be able to confidently extrapolate findings to the entire population prescribing, dispensing and taking the drug?"

And we are supposed to go to 11:30, which means about 35 minutes, or we cut into lunch. So let's start and see where we can get on this question. Who wants to start off the discussion? Dr. Krotki has his hand up and then Dr. Lee and Dr. Lemeshow. Okay.

PANEL MEMBER KROTKI: I'm not sure these are the right questions. The main question that I heard, that I drew from everything I have heard so far is the frame. Do we have a complete and accurate frame?

Once you have a complete and accurate frame, or at least you know what its deficiencies are, then you can draw -- drawing a sample from a frame is not rocket science, and neither is determining the sample size. I don't see those as challenges. I think the main challenges are

the frame.

MODERATOR TOIGO: Do one of our panel, one of our FDA folks want to comment on that?

DR. IZEM: Yes, this is Rima Izem. If I could have slide number five please. With regard to sample size, I just wanted to make just a comment. It's not as trivial as you may think.

For instance, here we are considering different paradigms, two different paradigms here, one that may be more familiar for reviewers like me who review consumer behavior study is the hypothesis testing paradigm, that is based, you have sample size of your -- of the sample of your survey, on the target rate, and that's sort of the driving principle of the target rate.

Once you specify where that target rate for the primary endpoint is, then you can derive what sample size is appropriate

for your survey.

The other driving principle or the second paradigm is the precision of the estimate. And that's the one that was listed in Dr. Shibuya's slides this morning that is sometimes used by applicants.

And that paradigm may be more familiar, actually, to people who are conducting national surveys. However, it's less familiar to people like me who review consumer behavior studies, because that's not the paradigm that is used there.

So Dr. Shiffman mentioned a lot of the consumer behavior studies. Those are label comprehension studies, actual use studies, and self-selection studies, and the driving principle there is hypothesis testing.

So those are the two different paradigms they would like the panel to weigh. Thank you.

MODERATOR TOIGO: Dr. Lee did you

want to start us on that discussion, or did you have another point you wanted to raise?

PANEL MEMBER SUNGHEE LEE: It's sort of related.

MODERATOR TOIGO: Okay good.

PANEL MEMBER SUNGHEE LEE: So it seems like the target population both prescribers and the patients, it seems like they are sort of a moving target. It seems like they change. As you have more patients taking the medicine they will change over time.

So I was wondering how they will be defined to begin with, how the target population will be defined to begin with, and also is there any viable way for suggestions or whatnot to estimate the total -- estimate the target population size? I think that will sort of determine your sample size as well, because if you are talking about really small, I guess, set of patients taking the medicine as the target population, then you

know, your target sample size will be much smaller than medicine that will be taken by the broader population.

MODERATOR TOIGO: Did you want to comment on that Rima?

DR. IZEM: Yes, the two paradigms, obviously you'll have different sample size calculation if you have a finite population frame. So those are general paradigms. We are not looking for specific formulas. Obviously the formulas for each of the paradigm will vary depending on your sampling frame.

So if you have a stratified sample, cluster sample or a finite population sample that will affect obviously the sample size calculation.

But at least the driving principle is very different for these two paradigms. One would be focused on the target rate and the other one would be focused on the precision of the estimate.

Does that answer your question?

MODERATOR TOIGO: Dr. Lemeshow.

PANEL MEMBER LEMESHOW: I just wanted to comment on, in question 1A, your question was is it possible to generalize from a sample to a population, the answer is of course. All the sampling theory, there is a large sampling theory that tells you how to generalize from samples to populations.

The important point there is the word representative. I mean, the sample has to be representative of the population in order to be able to make that kind of inference.

Now, sample size calculations are very different in situations where you are interested in hypothesis testing versus estimation. You can easily set this up either way. You could say I want to estimate the proportion of the population who receives this drug who understands the risks. That's an estimation problem.

The hypothesis testing problem is that we think that given that they get this information, that the proportion who should understand the risks correctly is 80 percent and then you can test the hypothesis that in fact it's 80 percent. But the sample size calculations are totally different in those two situations and I think you have to determine which of those you are interested in.

And finally, there are -- the issue of whether populations achieved certain goals in terms of percentages is something people have considered for a long time.

I mean, maybe some of you know that in work in developing countries, a question that's long been asked is has the World Health Organization achieved its goals that 80 percent of children are vaccinated for childhood diseases?

So there's a goal. Eighty percent should be vaccinated, and they do surveys

that are very different than the ones that most of us think about.

For example there's a survey technique called lot quality assurance sampling and it's set up in such a way that you are told to take a random sample -- now that's no minor task -- to take a random sample of the population and then if more than, and say a random sample, and the number that you take is also fixed, so let's say you take a random sample of 12 people, if more than 3 of them don't know the answer, this fails the test, and then you have to some kind of remediation.

So -- but if more than, you know, 10 out of the 12 understand the question, then you have satisfied the goals. Now, in one sense that's the simplest possible way of asking the question: do you, have you satisfied the goals set out by the FDA, yes or no?

And if the answer is no, then you

do something. If the answer is yes, you move on. And there are a lot of different sampling methods.

The other point was establishing a frame that I wanted to say. When -- it's not necessary to establish a frame for every single person prescribed a drug. But what is necessary is you understand what the general population is, and then you can work your way down by stratification, clustering, whatever, until you get to a smaller level, and then you construct the frame, and then you can take your sample from that smaller subset, and then you build up a sampling frame and they you can do your estimation.

MODERATOR TOIGO: Thank you. Dr. Raghunathan and then Dr. Gilsenan.

PANEL MEMBER RAGHUNATHAN: Yes, I think that going back to the frame construction, I think at least we have pharmacy chains that dispense the drugs that could be sort of a primary sampling unit, so

you could sample, we have a collection of pharmacy chains and then that could be, you form that primary sampling units, and then you can take the patients within the -- but within those primary sampling units to create the representative sample that way.

And also healthcare providers, there must be a list of all the dispensing doctors or prescribers, and there must be a list of people who belong to hospitals or belong to HMOs and there could be multiple levels of clustering that we could use in order to sample -- in order to sample a representative sample from the population.

MODERATOR TOIGO: Rima, you wanted to comment on that?

DR. IZEM: Yes, I would like just to ask for a clarification. By sample do you mean randomized, a random sample? Or do you mean just a subset of this frame?

PANEL MEMBER RAGHUNATHAN: I would say that I was meaning taking a probability

sample of pharmacy chains, a probability sample of healthcare providers. I think that that's critical.

MODERATOR TOIGO: To that point, Dr. Morris or -- okay.

PANEL MEMBER MORRIS: I think we are missing the practical applications here. I mean, sure, in theory you can do all of this stuff. The reality of it is you can't take the random sample of pharmacies because pharmacists won't agree to do the survey, and there's real practical limitations to finding these people, to getting them do to the surveys.

I think we are talking about the head of a pin here. This isn't realistic.

MODERATOR TOIGO: Thank you. Dr. Gilsenan and then Dr. Shiffman.

PANEL MEMBER GILSENAN: So there's a couple of points in response. For sampling patients, typically we do actually go through a pharmacy chain to get the patients and

logistics is the key because you have to go with someone that's willing to participate and so therefore it's a subset of the pharmacies that you are starting with, and then it may be a subset within that subset that you can work with.

But typically, you can get a list of patients who have been prescribed the drug within a certain time frame and then you can randomly sample from that population or in some cases, you have to take all comers because there's just not enough of the -- there's not enough to sample.

So that's often what we do for drugs that are distributed through retail pharmacies. And then also with the physician surveys, we do exactly as you mentioned, we get a list of prescribers from the sponsor and then -- and then we sample from that list.

One of, I think one of the issues is once you have the -- you get this list and

there's a lot of restrictions on the amount of detail that you can have in that list with regard to your sample, but a lot of it is due to patient privacy or willingness to participate and have your name put on the list.

So it makes it very challenging but I think it is important to do some sort of assessment of the respondent and non-respondent populations, so once you have your list there is very little information you can get from these pharmacy chains to do a comparison. You can usually get their age and their gender and whether it was a new or a refill prescription in the database, whether that's important or not, the type of insurance they have.

And then you can do some comparison and that will kind of tell you if you're in the ballpark or not. And then with the physician information it's quite limited as well and geographic distribution is another

thing that you can look at.

Another thing that I think would be very helpful to get at and it's not as easy as you would think, is for the sponsor to provide information on their patient population, like what's the general age category, what's the gender distribution, and other things.

It's very difficult, as was mentioned earlier, to do that when you have a new drug on the market and those things change over time.

But I do think, even though there is limited information available, attempts should be made to try to look at the response and non-response even with the practical limitations.

MODERATOR TOIGO: Thank you. Dr. Shiffman.

PANEL MEMBER SHIFFMAN: Saul Shiffman. I think this will build on comments that have been made. I mean, the

small samples you are talking about are the small samples of completers. So the issue hasn't been that people aren't doing reasonable sampling in terms of sending out the instrument.

So there are two -- there are really two reasons you could have a small achieved sample. One which I think we need to talk about is the entire population may be quite small. Some of these drugs, whether designated as orphan drugs or not, may have a very small number of patients using them and I guess I'm going to pose a question to the statisticians on the panel about whether there are different methodologies when you have a small sample but it's actually quite a large proportion of the population.

But the major reason for small samples is non-response, and I would just echo the comments that have been made that that's of concern not because of the size of the sample, but because of the potential for

bias, and that very often these samples don't allow enough information about non-respondents.

So I think the issue is less what is the size of the sample than what do we know about the potential for bias and how could we be informed if we understood more about the non-responders, and specifically with regard to physicians, we need to distinguish a variety of reasons for non-response and something that we haven't touched on at all is that whereas a single patient with a particular condition and drug may just ask to be asked to fill out one of these, with the increased number of REMS, an individual physician may have 10 REMS in which they are participating.

And I think part of the non-response is just that the cumulative burden, not the single REMS burden, becomes too much for physicians, and that suggests a need for coordination that isn't currently present.

MODERATOR TOIGO: Thank you Dr. Shiffman. Dr. Willis, Dr. Lee.

PANEL MEMBER WILLIS: Okay, I think I have a radical perspective on all this, but we all work with very limited resources that we have to use in particular places, and not elsewhere.

And I wonder how realistic it is to think that we can do statistically valid estimates of knowledge rate, especially for small populations, given all the sources of error that we observe. I mean, how much confidence are we going to have in a 70 percent, 80 percent rate with a certain standard error?

I mean, where I work we do huge, huge tobacco surveys, and our large surveys don't even agree with respect to what the smoking rate is in U.S. populations.

So if we can't get stable estimates, how are you going to do it for a REM here where you have a small population

and maybe a biased sample, high non-response rate.

As a qualitative researcher, I would rather in -- I would advocate putting resources into our developmental, pre-testing, cognitive interviewing types of activities to try to produce patient education or other materials that are likely to be understood in the first place, like, that's where the pharmaceuticals I think can get the most bang from the buck, personally, is development rather than putting everything into the assessment to produce metrics that I really would not have much confidence in in the end.

MODERATOR TOIGO: Thank you. Dr. Lee.

PANEL MEMBER SUNGHEE LEE: Kind of to respond to -- I didn't get your name, sorry -- the proportion of the sample from your population, if the proportion isn't really large, it'll affect your precision

because you will have to use a finite population correction factor, and that will decrease your sample error.

And in terms of non-response bias, and also it can be probably bias from non-, probably sampling as well, I saw, I think it was some sort of a physician drug in diagnosis audit data, is there can be a source to test biases in your study.

MODERATOR TOIGO: Do you want to -- were you going to comment, Dr. Lee, on Dr. Lee's comment?

PANEL MEMBER MATTHEW LEE: Yes.

MODERATOR TOIGO: Go ahead.

PANEL MEMBER MATTHEW LEE: This is actually to Dr. Lee and Dr. Shiffman. I think respondent burnout is a huge problem. It's going to become a larger problem and enter more bias into these REMS surveys as they move forward.

I mean, when you think about class-wide REMS, let's take anti-epileptics for

example, you've got a dozen different agents that are probably being prescribed by the same prescribers. You have surveys that are going out not just at 18 months, 3 years and 7 years, but you have some that are going out every 6 months, every 12 months, for all these products. You are eventually going to dwindle down the number of respondents that you are going to have. You are going to have a larger number of bias moving forward, you know especially as more products come into the marketplace too. You're just going to -- you know, you've got a very limited resource as far as respondents, especially when you are dealing about prescribers and not even taking into consideration orphan drugs or other conditions whether there's an even smaller number of prescribers, how valid are the results going to be seven years from launch of your drug?

MODERATOR TOIGO: Thank you. And that's why we'll talk about alternatives this

afternoon I guess. Okay, who's -- Dr. Chowdhury and then Dr. Raghunathan, did you have your hand up as well? Okay. Okay.

PANEL MEMBER CHOWDHURY: Having a good and complete frame is important for many different reasons. If we have a good frame, only in that case we can measure the accuracy of the estimate, we can draw inference from this if we have a good, then if we know the size of the frame and distribution of the frame, then we can also utilize this in making some adjustment for non-response.

So the whole thing I guess, it's very important to have a good and complete frame, but in a practical situation, having a good and complete frame may be very difficult in many situations. But there can be a difference between the inverse and the frame, a slight difference and in some cases considerable difference.

If even the inverse is very large, but if you can identify a large number of

chain of pharmacies, and then we assume that there is going to represent sort of the universe, then we can consider that as a frame and use this frame to draw our sample and then relate all the inference and everything to this frame, and that frame may be slightly different than the universe.

MODERATOR TOIGO: Thank you. Dr. Raghunathan.

PANEL MEMBER RAGHUNATHAN: Yes, going back to this notion of representativeness, representativeness os assured through a proper stratification of the sampling frame and then drawing a probability sample from that properly stratified clustered frame.

So I think that just taking anybody who responds to the survey and then saying that how do I make it representative, is completely non-starter.

So I think one has to really make an effort at least. I don't think it's

difficult but I think serious problems require serious solutions.

So at least you have to make an effort in terms of creating a frame, and then attempting to draw in the sample from that frame using the probability sample method, assess the response properties and assess the non-response properties, and that's -- if you really want to do a serious job of estimation, that's what, or you might say that well, we are going to really not really do any kind of a survey, just go do a qualitative assessment and that's also an alternative way. But I think if you are really wanting to answer these questions in terms of representativeness, then I think one has to really seriously construct the frame.

MODERATOR TOIGO: Rima, were you going to comment there?

DR. IZEM: Yes, I would like to add another ingredient, I guess, to the conversation of representativeness. So there

are variables that you collect on the frames or maybe on the participants that could relate to knowledge, and there are some variables that do not relate to knowledge.

And obviously, representativeness may have a different results -- consequences on the results that you get whether or not they are related to knowledge.

So I just wanted to add that sort of comment to the -- ingredients, I guess, to the discussion. For instance, since so many people have brought up the actual use studies or the consumer behavior studies for over the counter drugs, for label comprehension studies for instance, the samples that are used in the label comprehension study is not a random sample. It's actually an all comers sample. However there are conditions or there are conditions that have to be met, which is that the sample has to be representative of the literacy rate in the U.S. population, because it's considered that literacy will

affect comprehension maybe much more than other variables. So I just wanted to bring that to the discussion.

MODERATOR TOIGO: Dr. LaVange did you want to add to this or a new topic?

PANEL MEMBER LAVANGE: Yes, just real quick. So just to second Dr. Raghunathan, that if you can draw a random sample, if you have a good frame and you can draw a random sample, the beauty is that you will get representative for the variables that you don't know about, or that you are not available.

So you can maybe work in a purposive sample if you know every single variable that is going to affect response. But we like to fall back on the random sample because there's variables we don't know about that affect knowledge or that affect comprehension or affect attitude, and that's the only way we could ever have a chance of saying something is representative. So I

just wanted to put that plug in.

MODERATOR TOIGO: Okay, and Dr. Stemhagen on this point as well.

PANEL MEMBER STEMHAGEN: I'd just like to echo Dr. Morris, in terms of practicality, because there's a lot, especially with patients, that HIPAA is very important, and so for sponsors to get a lot of information about patients before they have agreed to opt in is not always possible.

And so there are some pharmacies that agree to that, whether that's a representative sample of pharmacies, we work with nationwide sets of pharmacies and not only certain ones will agree to even give you information on a patient, but they need to send out the invitations and have the person opt in.

So there's a lot of confidentiality. They certainly don't have a lot of variables that you are mentioning. We can collect them in the survey. A survey for

instance with literacy that we are talking about, education levels are surrogate, you know, when we do our qualitative testing, we do the realm and we do other things on our materials, but that's not practical in terms of an actual survey to do things like that.

So we have to figure out some surrogates and then what's the appropriate population. Is a national sample appropriate? We are talking about patients with a certain disease, and so are they different than the general population? Probably they are.

And then to get that kind of general information on patients with MS or patients with epilepsy or whatever it is, those kind of national data don't always exist.

So I just want to -- there's a lot of, again, of practicality that has to come into how you are going to select a sample that you think is going to be the best

representative sample that you can. You know, if you have a REMS where you have enrolled everyone, that makes it a lot simpler. You know the sampling frame, you have some demographic and other characteristics. But those number of elements to assure safe use REMS are small compared to what we have been talking about although a lot of the Med Guide ones are going away, that you know, you have a lot more chance of getting a better sample when you have everybody registered is the obvious point.

MODERATOR TOIGO: Okay, I have three names on the list, Dr. Cantor, Dr. Krotki and Dr. Morris. Are any of those relevant to this particular topic?

Okay, well then we'll -- to this? Yes, okay.

PANEL MEMBER HOLMES: I'll just add, another practical issue is even if you can draw a sampling frame and draw a sample

that you believe is representative, in some of these cases, just to draw connections between a couple of folks and Dr. Lee's point that a burden, particularly with HCPs, remuneration is going to be the key, and we have limits as sponsors as to what we can do in terms of remuneration.

So I just wanted to add to Annette's comment with regard to patients, that on the HCP side, we are going to be coming up against the burden and the inability to provide the kind of remuneration that we would need to do in order to get those representative samples.

MODERATOR TOIGO: Okay. Thank you. Dr. Cantor.

PANEL MEMBER CANTOR: Yes, and this might be anticipating, I guess, the discussion at 1:30, I'm not sure, but drawing the sample is one thing, but it seems like we are not talking very much about the process used to get the responses, which -- and the

constraints around those processes.

So what kind of -- what sort of mailings are people doing? Are they doing multiple mailings? Certainly we know a lot about how to get -- or to maximize response rates and how to minimize non-response bias. We do know that non-response bias is highly related to saliency, so if you are getting a very low response rate, it's very likely that you are just getting those people who are very concerned about it, and the implications of that may have to do with trying to standardize the procedures people are supposed to be using.

Another question I had was, what is the time frame -- it's still not clear to me what time frame the companies have to implement this survey. I have heard referenced the 18 months, 3 years, 5 years, it's -- not being in this in a very detailed way, it's not clear to me what -- at what point does the survey start and how long do

they actually have to do the survey?

That has huge impacts on the potential for non-response bias.

MODERATOR TOIGO: Dr. Willy, did you want to clarify that, comment on that? Or Brian?

MR. GORDON: The timetable for assessments is set in the REMS, and what -- the most common one is 18 months, 3 years and 7 years, and that means at 18 months they have to submit an assessment report. So they will have to design the survey, conduct the survey, analyze the results and then deliver the report at 18 months, then another one at 3 years, then at 7 years.

MODERATOR TOIGO: And Brian, is that in the statute, or that -- I think that would be helpful for people who don't understand that. We didn't make it up.

MR. GORDON: Right. It's in the FDAAA language.

PANEL MEMBER CANTOR: But so when

are they actually able to start doing the survey? Is it at month one or is it -- does the drug have to be deployed for a certain amount of period and, you know -- so can you standardize how long the survey should be in the field for example? And how many contacts they should be making to try to get responses?

MR. GORDON: We don't specify. We let the sponsors design their programs and we recognize that the REMS are -- they vary so depending on the drug and the population, or sometimes the timetables are different than 18 months, 3 years and 7 years, so it's up to the sponsor to design how they are going to implement their surveys.

MODERATOR TOIGO: Okay, I think -- are these comments relevant to that? I see the sponsors' hands going up, so let me -- let me give you an opportunity to comment.

PANEL MEMBER HOLMES: Two quick things. One is that even though there's the

18-month, 3- and 7-year, it's not infrequent, and I believe all the examples that the panel has gotten with regard to a lot of our medicines are all annual assessments.

And the second point is, I believe I am correct on this, that there is a requirement for last subject in of any REMS assessment, that that can't -- Annette can speak to that, but --

PANEL MEMBER STEMHAGEN: No -- our interpretation, no older than 60 days, so typically you need that 60 days when you finish your survey to do your analysis, write your report and so on. So usually coming up at 16 months, when you need to finish, because the FDA doesn't want data that are too old, that doesn't mean everybody is surveyed at 16 months. It means you started your survey at maybe 12, 13, 14, 15, 16, depending on what you think your sample is, how long it's going to take you to accrue that number.

MODERATOR TOIGO: Dr. Peterson.

PANEL MEMBER PETERSON: Just another attempt to clarify. As always, the answer is "it depends." But in general, I'll just take the example of an 18-month assessment. That, again, is when the report is due to the FDA. A key word came up earlier, and that's the word "enforceable," so that is not a deadline that we will ever miss as a sponsor.

Typically, these days REMS get approved when the compound gets approved. It's not always the case. So it does take a little bit of time before any patient is using your drug. The day of approval isn't necessarily the day it's on the market.

In addition, FDA requires that we prepare a survey protocol, submit it to them for review, and they get 90 days to review it. So in general, we can't start for several months nor would we want to.

So in the field we have, I would

say, at best about a year for an 18-month assessment, and then when you put it in depends on how long you think it's going to take to find the patients and do it.

But an important point is we do have REMS with 12-month due dates, so and then, we can be down to three, four months to have a survey in the field.

PANEL MEMBER CANTOR: And my understanding is that's sort of a rolling admission, so as people get enrolled they get a survey?

PANEL MEMBER PETERSON: Yes, that is true. Another challenge is on the other end, that requirement that from the day the last patient completes the survey, we have 60 days to submit the report. That can be challenging if there's other elements of the assessment that are due in that report. We have actually held open surveys to make sure we don't finish before that 60 day time frame, to coordinate it with other parts of

the assessment.

PANEL MEMBER CANTOR: Just one final question. I guess there was a slide up there about remuneration. So are they allowed to slip two dollars into the envelope to get people to at least pay attention to the survey request?

MODERATOR TOIGO: Do you want to comment on how you do your --

PANEL MEMBER STEMHAGEN: Typically you have to complete the survey in order to be paid. That's usually what happens for healthcare providers and patients.

So rather than sort of the enticement of sending the invitation with some payment, it's after the surveys are completed, and we are talking about, for patients, usually, these are 20- to 30-minute surveys, often, between 15, 20, $25 for the patient, and physician, has to be fair market value, based on the physician's -- the medical speciality and so on.

212

So that's evaluated based on some independent measures of fair market value and it's usually somewhere between $100 and $125.

PANEL MEMBER CANTOR: Yes, just one final comment. We know from a vast amount of literature that at least putting that $2 in, even though they may not return the survey, works -- is much more cost-efficient than asking people to fill it out on a promise.

MODERATOR TOIGO: Thank you. Dr. Krotki. Oh, wait, can you hold one minute so that Dr. Lee can add to this discussion?

PANEL MEMBER MATTHEW LEE: And this is very related to remuneration, and survey representativeness. There are some healthcare systems that don't allow the sponsors to even offer the practitioners or the patients any sort of remuneration for surveys or participation, thus biasing the sample, because they are excluded from us even being able to offer it or offer a donation to a charitable cause in the

physician's name, et cetera.

MODERATOR TOIGO: Thank you. It's always good to hear more of the practical challenges in implementation. So, Dr. Krotki.

PANEL MEMBER KROTKI: Before we become too despondent about all the challenges facing us, let me reassure everybody that taking good surveys is very difficult, not just in this context, but in the context of household surveys and many other establishment surveys.

There's all the issues that have been discussed in the last half hour are faced on a daily basis by people like myself who work in the survey research world, whether it's incomplete frame, whether it's non-response, whether it's non-response bias, whether it's measurement error, et cetera, et cetera.

So it would take far longer than we have at this table today to discuss all the

problems and potential solutions. I encourage people to look at some of the work, good work done by the members of the American Association of Public Opinion Research. There's actually a conference next week on -- the fourth international conference on establishment surveys, where there will be, I'm sure, a lot of papers on how to reach and get good surveys with physicians, hospitals and such establishments.

So, yes, so there are -- we are all grappling with these issues. Some of us who work on household surveys are facing 20 percent, 10 percent response rates, and yet we are still producing data and estimates that people think are credible.

MODERATOR TOIGO: I would ask that if there are some specific ones that you think are especially relevant for us to look at and consider as we grapple with these issues, that you could let us know and we will put those in the docket, or just give us

the references and help us comb through them if you have already identified them.

Okay, Dr. Morris.

PANEL MEMBER MORRIS: I just wanted to reiterate what Dr. Willis was saying, and it seems a lot of the discussion that we have been having is how much effort to we put into getting a precise estimate, if we are in a population.

And I always thought the answer to that question always went to how are we going to use the information. And my concept of how a REMS is set up, is that you are trying to basically have a quality control mechanism in that you are serving people over time and what you want to see is improvement over time.

And, yes, there may be a lot of systematic biases that go into any single survey, but if you are repeating the survey over time, in theory those systematic biases should be fairly constant and, in theory, you

are measuring some kind of gain in behavior.

So given the concept of a REMS, yes, you do want to get the best population estimate you can, but in terms of where I put my resources, I would go back to what Dr. Willis said, and let's make sure we are doing a good job in the first place and then try to build on it, rather than put all our effort into getting precise estimates.

MODERATOR TOIGO: Thank you. Dr. Lemeshow, and we are at -- we are a little over 11:30 so just, after Dr. Lemeshow then it's comments that people think are really important for us to have for this session

PANEL MEMBER LEMESHOW: So I don't want to be a curmudgeon here, but if -- for me, if the sample is not representative of some -- to me, taking a grab sample to get some information provides you no information about the population. I'd rather see no survey done than to have a survey done where you don't have representative, in some way.

I understand, there were big issues here in getting representative samples, but we don't learn a thing from samples that are not -- and if the sample done at time A is not representative of the population, the one at time B is not going to be representative either and you're not going to learn anything comparing A to B.

PANEL MEMBER MORRIS: Well, it's representative to some extent.

PANEL MEMBER LEMESHOW: It's not -- it's not -- taking a -- look, if I'm standing outside of a grocery store and I want to learn something about people's buying behavior for cereal, and if a guy is coming to the grocery store with a baseball bat because he's just mad because his wife just got poisoned from something he bought at the grocery store, you're not going to stop this guy and ask him what do you think about cereal?

And that guy's opinion is very

important because he is part of the -- he's part of the population you're trying to understand.

And if you simply take who you get from these surveys, you will learn nothing about the population of interest. If that's -- if it's not of interest to learn about the population, then you don't have to take a survey.

PANEL MEMBER MORRIS: But you're always getting who you get, in any survey.

PANEL MEMBER LEMESHOW: That's not true.

PANEL MEMBER MORRIS: There's no such thing as non-response?

PANEL MEMBER LEMESHOW: Of course there's non-response but there's --

PANEL MEMBER MORRIS: So you're always getting who you get.

PANEL MEMBER LEMESHOW: But it goes in part of the survey design. You design your survey --

PANEL MEMBER MORRIS: Well, you want to do the --

MODERATOR TOIGO: So maybe this will be the lunchtime discussion between Dr. Morris and Dr. Lemeshow, and -- because unless someone else here wants to jump in, join the table for lunch.

So I think Dr. Shibuya do you have enough on this question for this morning? Okay, so we'll break for lunch. You've got one hour and we'll be back at 12:35.

(Whereupon, the above-entitled matter went off the record at 11:33 a.m. and resumed at 12:35 p.m.)

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

(12:35 p.m.)

MODERATOR TOIGO: Good afternoon and welcome back. For those of you who are just joining us, I am Terry Toigo and I am moderating this meeting.

First, thank you to the panel members and to the speakers for this morning because you kept us -- you helped me keep this meeting on time, so I appreciate that.

A few housekeeping details, again for this afternoon, if you haven't already turned off your BlackBerrys and cell phones and other devices, if you could please do that.

And then for our afternoon session, the open public -- open public session speakers, if you haven't signed in at the desk, if you could please do so, we'll have a 15-minute break around 2 o'clock and then we will reorganize for our second panel and our open public session during the break.

And if you remember from this morning, I got a question about comments or questions from the audience and how we were going to deal with that, and as you can see we are using all the time allotted.

At the same time, we do -- this is an open process and we would like to accommodate your questions. So if you have questions that you think we should have addressed at this workshop, or that the panel should have addressed, if you can take those to Julia, if you can write them down, and write down your name and your affiliation, and if you can leave them with Julia and I will try and before the next -- before we open the meeting after the break, I will read the comment -- the questions into the docket, or into the record, assuming that we have a manageable number of questions that need to be taken care of.

So we are continuing with panel 1 and we are on question 3, and we are at a

point where we are using the right number, so we are not juggling the number, so this really is question 3 in your handout materials, and it's on questionnaire design.

And I'll read the question for the record: "Since most surveys use only true/false and multiple-choice questions, what are the advantages and disadvantages of using other question types, open-ended, case vignettes, fill-in-the-blanks, to evaluate knowledge?"

And we'll see who wants to get us started for the afternoon session, with responding to that question.

Nobody wants to start? So perhaps one of our FDA panel members here wants to put -- provide a little more clarification on that question, or kind of tell us exactly what you are looking for that you want the panel to maybe grapple with.

You don't want to leave that to me.

ASSOCIATE DIRECTOR WILLY: Okay,

this is Mary Willy, and I guess we are looking at just getting -- having a discussion and maybe it's -- we don't need the full time, but about the options for using these different strategies for questioning folks who are taking the surveys.

So I think we already heard someone speak to true and false, but there are other options using case vignettes and so forth. So if people would like to comment on that, this also gets to when we are talking about questions, how many questions should be included, whether it's good to have repeated questions to some extent, and if so then the question of how many questions is realistic and reasonable.

MODERATOR TOIGO: Dr. Willis.

PANEL MEMBER WILLIS: Okay, well, to start, classically the advantage to an open-ended question is that it's unbiasing. I mean, it's not leading in any ways. The answer has to come out of your head.

The disadvantage, of course, is they are very difficult. I mean, what are the risks of a certain drug, well, I have to come up with them out of my head or I can look in the materials depending on whether this is an open or a closed book exam, I guess.

So we typically say, oh, that's too hard, let's go with another format, true/false or a select from a multiple list, and there the advantage is, well, recognition memory is easier than recall from a cognitive point of view.

The real disadvantage to that in this kind of environment is it becomes really important that what those distractors are, that you put as the incorrect responses, and that's where it becomes a game, you know, any kind of test like this is a game, and any game can be gamed by either the respondents or the test designer.

And so it's easy to see how a test

designer could decide, perhaps, if they want to reach that 80 percent of other magical level by selecting distractors such that it becomes a really easy test, and someone else might be penalized because they come up with a much harder test and they only get a knowledge score of 60 percent, and yet the difference is an artifact of the production of the instrument rather than of anything coming out of the respondent's head. So that's just an issue for consideration.

MODERATOR TOIGO: Thank you. Dr. Stemhagen.

PANEL MEMBER STEMHAGEN: Just one comment in terms of the practicality, things like case vignettes are very powerful and can be very useful for behaviors and so on, but when we try to -- when we are designing them think about the fact that we want to offer multiple ways for respondents to answer, so it may be the internet but it might be phone interviews and a lot of that, if it's a very

long vignette, is very difficult to do on a phone interview and have people understand it and make sure.

So that's certainly something that we always think about when we are trying to design whatever type of question it is. How is going to work on the phone as well as how is it going to work on the internet?

MODERATOR TOIGO: Thank you. Dr. Ostrove.

PANEL MEMBER OSTROVE: Nancy Ostrove. I hadn't really thought about the phone thing, but that is definitely an issue that we have run into in the past in constructing phone questionnaires.

But on the other hand, I mean there is a real, I believe that there is a real positive aspect to using like a case vignette, especially -- not necessarily for the patients but, well, I take that back, for both populations, for both healthcare providers and patients, because what that

gets you to is, is again, getting kind of a better sense of the person's real in-depth knowledge, can they apply the information to a situation that goes beyond just the verbatim parroting back of a particular piece of information?

So, you know, I realize that there might be issues, and administering it over the phone is just a piece of it because once you go to more in-depth knowledge measurement, then for instance, if you go to open-ended questions, then you've got to deal with coding and you've got to deal with reliability and the resources that are required for that and the time it takes to code and then analyze that becomes much more problematic for the manufacturers, obviously, and it just adds to the time and you've got all these time issues.

But vignettes you can -- you can probably do more with some closed-ended responses so you don't have that part of it,

so there -- I think you have to look at the tradeoffs with respect to phone administration and given that I've heard a couple of people talking about using mixed models, that might be another way to do it, mixed modes, excuse me, of administering the survey.

You know, I would say that if you really want to get to the in-depth knowledge, to the application of the information, that people ought to really consider very seriously going to kind of case studies and looking at those.

MODERATOR TOIGO: Thank you. Dr. Raghunathan.

PANEL MEMBER RAGHUNATHAN: Yes, I think before we talk about the questionnaire design I think one should be clear about what are the important information that you expect the -- either the healthcare provider or the patient is expected to have, based on the material or based on absolute knowledge.

But then, you know, then there's no question of either/or, when you use case vignettes, talking about the particular situation, and then ask based on that vignette, you know, Likert scale, somewhat likely, less likely to prescribe or more likely, somewhat less likely to prescribe, or those kind of five point scale, four point scale, or true or false question could be asked.

But the case vignettes kind of gives you a proper context by there, a person can be really primed to understand what the context is and then based on that context, what the understanding is of that -- for the patient or for that healthcare provider.

So, first I think we should really discuss about what is exactly the knowledge or awareness that they are -- they should be possessing, and then develop a case vignette for each of those, measuring each of those knowledge awareness.

MODERATOR TOIGO: Thank you. Doctor, let's see, Gilsenan was next and then Dr. Morris.

PANEL MEMBER GILSENAN: Alicia Gilsenan from RTI Health Solutions. Just in our experience, the underlying structure of questionnaire development has to be used and guide what -- how you construct the true/false and multiple choice questions and sometimes it's a little more practical in these surveys.

We have done case vignettes typically with the provider surveys but you have to be careful with the true/false especially because it can be very leading and someone could guess the right answer very easily without ever having been aware of the information that you are after.

So I think, critical to the types of item responses is also the testing that you do ahead of time to make sure that you are assessing what you think you are

assessing in a way.

And then as mentioned before, the more complex response choices you do have to take into consideration the analysis and the time needed and the interpretation when you get that type of information.

MODERATOR TOIGO: Thank you. Dr. Morris.

PANEL MEMBER MORRIS: Yes, I think a lot of this goes back to what you are -- what you mean by knowledge and I think there's been a lot of discussion this morning about knowledge is actually lots of different things.

And I guess I always like to discriminate between surface knowledge and gist knowledge, and people are going to read a document and understand, you know, what it says initially, but then over time they'll forget the details but hopefully retain the gist.

And I think that -- I would call

them scenarios rather than case vignettes because -- but some question that measures, how people apply the information, how they make decisions, I think is helpful to supplement more traditional forms of knowledge because it gets you a little bit more into application and decision-making, so you know, you have to choose, which is your primary measure, but then these other measures might, you know, help you to understand and elucidate what you are actually measuring, what people really know.

MODERATOR TOIGO: Thank you. Dr. Shiffman.

PANEL MEMBER SHIFFMAN: I guess one important principle is that you want your data collection to be a test of the knowledge, not a test of the test, which is to say that you need to be very sure that the test structure isn't what's creating, especially incorrect responses.

And let me give a couple of

1 particular examples, which from our
2 experience do very poorly, keeping in mind
3 that by and large these are self-administered
4 surveys and I have in mind here the patients
5 rather -- the physicians are really expert
6 test takers, so they can handle very complex
7 questions, not so of patients and consumers.
8 And so for example open-ended data
9 whether it's from a vignette or other source
10 can be extremely valuable but it really
11 doesn't lend itself to self-administration
12 because even before you get to the coding
13 people have to produce enough information to
14 be codeable.
15 And so when you do self-
16 administered, open ends, people often give
17 you a response which is basically
18 unresponsive. It is not that it shows they
19 don't know, it doesn't show anything, because
20 they will say something irrelevant.
21 So I think open ends simply can't
22 be used in a self-administered format. You

need to have an interviewer who can, in a non-biased way, probe.

But another example of the multiple choice questions where more than one can be true or many can be true, again, physicians have seen those on boards. They know how to do them. Consumers don't.

And so that's an example where the item ends up structuring the responses rather than the knowledge, so I think the basic principle is the test structure should be familiar and comfortable to respondents who are not good test takers and should be clear enough that you know you are testing their knowledge and not tripping them up with a complex question.

MODERATOR TOIGO: Thank you. Dr. Ellenberg.

PANEL MEMBER ELLENBERG: A question for edification. I'm warming to the case vignette approach that has been mentioned by many of the advisers around the table, for

the reason that it captures more the functional aspect of what you do with the knowledge.

And my question is when you send out a case vignette, how do you make sure that there's only one answer? You will give a, as I saw the case vignettes before, you will give an example with four or five conditions and then say what would you do. But what if the specifics of a case that the physician, if it's the physician questionnaire is answering, is considering, he thinks of something else.

So he can't give a single answer to that. Do we know, do we have experience in how to get rid of that problem? I don't know. So --

MODERATOR TOIGO: Does anyone around the -- any of our panelists want to comment on that? Okay, Dr. Holmes and then Dr. Beatty after Dr. Holmes.

PANEL MEMBER HOLMES: So, I agree

that I really like these case scenario -- these case vignette and scenario ideas for the reasons just mentioned. And our own experiences with healthcare providers in particular, we often need to specify over and over again in our item stem that we are asking them to respond to what the prescribing information says because healthcare providers bring their own experience and their own biased views and their own sense of what the literature shows and how broad and complicated that can sometimes be.

So I think, just buttressing what you were just mentioning, Dr. Ellenberg, is that I think case vignettes may be a particular problem for healthcare providers, because you will then be offering them, as opposed to limiting them with that based on the prescribing information, you are now going to be opening it up to them for how they think oftentimes, in a very complicated

way, this is a case, I have seen this case before, it worked well, this is what their -- I know this would work for that case better than it would for another and I'm going to ignore what the prescribing information says.

So I think case vignettes bring particular problems for healthcare providers and I am not sure there's a good answer to the question, and it may be better for consumers. I don't know that.

MODERATOR TOIGO: Okay, I see two hands I think. Dr. Beatty, is yours a new topic or -- okay, so then how about we do Dr. LaVange and Dr. Gilsenan?

PANEL MEMBER LAVANGE: I just had a quick comment based on that. Isn't that what we want to know? I mean we don't want to know just what the prescriber got from the educational material. We want to know if the prescriber is aware of the risk and takes that into account in that case.

I mean again, thinking about the

public health impact, I want to know -- I can't measure the action so I want to measure what happens just before the action, and I don't really care where it came from.

I mean, I do care, but you know, you --

PANEL MEMBER HOLMES: Can I respond to that from the sponsor side, which is we really want to know what the FDA wants to know. We are getting the sense that they want to know what the knowledge is and a regurgitation of that knowledge by the prescriber.

So to build in a case vignette where we might actually capture how a provider actually takes all of that into account, isn't really what we are being asked to provide here, and perhaps exposes sponsors to risk if FDA really just wants to know, do they know what FDA wants them to know?

And so I think that's answer that FDA maybe needs to answer.

MODERATOR TOIGO: Okay, I was going to let our FDA panelists comment on that. But we will go to Dr. Raghunathan first.

PANEL MEMBER RAGHUNATHAN: I think that you combined two questions. So one is that whether the provider gets the knowledge based on the material that you have sent, versus the knowledge that he or she has based on the practice and literature including the material that you have sent.

So maybe we are to ask questions separately to tease out is that based on the material, what would you choose or what options are the right options, or -- and the based on your practice, then what are the -- so that way you are covering both aspects so you are to tease out that if you are really interested in attributing or adjudicating the knowledge gained from the material that the sponsor has sent versus the knowledge that they have or awareness they have --

PANEL MEMBER HOLMES: And that gets

back to, are we really measuring knowledge. So where knowledge, practice functional understanding. So the outcome is different in all those cases, correct?

MODERATOR TOIGO: On that point Dr. Shiffman? Okay.

PANEL MEMBER SHIFFMAN: Yes, Saul Shiffman. Well, I guess speaking maybe for FDA, it's important to realize FDA doesn't regulate the practice of medicine. So the doctor is entitled to act in ways that deviate from the label so it becomes a problem. You are then assessing something that is of interest, but not strictly within FDA or sponsor purview.

But I was also going to just note that in testing of OTC labels with consumers, it makes heavy use of scenarios, or vignettes, but there is a complication to that, which is to decide how much information to include, which is either distracting or telling.

So it's not uncommon for consumers to react to a vignette to say, "Well, but I don't know this other thing about the person. How old are they?" So that's where information is missing.

On the other hand, if you put in a lot of information that is irrelevant, the question just becomes muddled and confusing. So scenarios are very attractive from the perspective of practical decision-making, but they are actually very difficult to craft well, and again I think are particularly difficult in a self-administration context.

MODERATOR TOIGO: Thank you. Okay, yes.

PANEL MEMBER GILSENAN: I was just going to respond to the earlier question about how you handle the case vignettes just with an example that we have used with prescribers, is we allow the physicians to select all that apply.

We have a specific choice that we

are looking for based on the risk information but there may be other choices where the physician, you know, one example might be the person -- people over 80 years old, that's not relevant to the specific risk, but for that physician, he may decide it's not appropriate to prescribe to that person and so there's a right answer and a wrong answer within the choices and then there's some other things that really don't count against them in their response.

And then the other thing that is really important, we've said this over and over, is to cognitively test with physicians, especially on these vignettes, to make sure that you are getting across what you want, because the physicians will want to know more information and then you have to balance it with the burden of giving them too much.

MODERATOR TOIGO: Dr. Beatty.

PANEL MEMBER BEATTY: I almost forgot what I was going to say. I have had

thoughts about various points in this conversation, so this might be a little bit broad-ranging.

But I guess when looking at all the options you have for asking questions I kind of see them along a continuum, with maybe true/false being at one end of it and more open-ended things being at the other, and both extremes have I think rather significant disadvantages.

True/false, as some people have said already, you know, they are kind of, they are easy to guess and they are also easy to game. I don't think that they are -- it's a really strong kind of way to ask a question.

But open-ended questions, whereas it's the most pure way to see what can you recall about all this stuff, it has difficulties in terms of administration. It also creates a sort of a, I mean I know you are, in a way you are quizzing people, but

you don't want them to feel like they are being examined, it's just kind of -- it's a less pleasant, less, people might be less willing to do it, and might get more tired of it more quickly.

It also doesn't allow them to really save face if they don't know the answer. So the stuff in the middle is what's, I think, a little more interesting to me. And when people talk about multiple choice questions, I think there are really great ways to phrase multiple choice questions. One is when you sort of have a question and there are multiple things that could be true -- multiple things could be true of them. It's not necessarily just one of them, and you get to evaluate -- use what you know to pick which scenarios in there are correct or which statements are most correct, as opposed to others.

Well, I like that a lot. Vignettes are kind of intriguing also. They are not

245

without their problems. I have used vignettes in kind of other types of environments when it wasn't possible to -- I was interested in more than just regurgitation of what someone knew. I wanted to see if they could apply what they knew, or how they would apply what they knew.

And the vignette was a way to kind of create scenarios that wouldn't have occurred naturally by chance, that I wasn't likely to just come up with, so I could ask someone to react to particulars that I could script that were important, and if well chosen, that can be great.

The down side of vignettes though is that the judgements could be really, if you are not careful, artificial, and make -- ask people to make decisions that, or judgements that they really wouldn't be like the judgements they would have to make in real life. You are just kind of getting them to a story.

246

But it's an approach that -- and again the conversation has kind of gone on a bit since I started formulating these thoughts, but vignettes I think are worth thinking about seriously. It's just a matter of how you construct them. There are good ways and bad ways.

MODERATOR TOIGO: Thank you. Dr. Sleath, Betsy, yes.

PANEL MEMBER SLEATH: I've been called Stealth too. I just want to comment. We haven't talked about the fact that a lot of time s you want to use open-ended questions first to develop your response categories, and because a lot of the samples we got, you can tell are written by people who know what the answers are but if you use open-ended questions to generate your response categories, that might help you catch when people are misinterpreting things in ways that you never thought of. Does that make sense?

So a patient that is asked to respond open-endedly could come up with something that providers would have never thought of so that's one suggestion.

The other thing is that Meredith talked about the use of theoretical frameworks and models and for those of us in academia, if any of us are going to get grants funded, we have to use some type of theoretical perspective or model, and to me this work seems to be being done kind of in a vacuum, with no model and there's a lot of behavioral science literature out there that could help if you want to get at knowledge, attitudes, behavioral intention, for each of those elements, you might want to do focus groups and ask open-ended questions to get at how you are going to measure that for a particular disease state.

MODERATOR TOIGO: My thinking would be that the sponsors probably do a lot of that stuff when they are developing their

whatever approach they are going to use. No, they don't do that. Okay. There's not -- okay.

PANEL MEMBER SMITH: Some of them do. I think there's just huge variability out there. That's all. Some do and some don't and maybe the most don't, but --

MODERATOR TOIGO: Okay. We might not see it all but I am assuming some of this stuff goes on in developing this.

Okay. Dr. Holmes, did you want to comment?

PANEL MEMBER HOLMES: And to be fair, I mean, I think that the time lines are very short, and some of this qualitative work is repetitive, builds on itself, and can't be done in two days or a week or a month, and so I do think sponsors are doing this work, in curtailed fashion.

MODERATOR TOIGO: Okay. So it's the time lines really that are impacting doing more of it or better.

PANEL MEMBER HOLMES: And it's always a moving target also for all the reasons we have talked about, sometimes early in a life cycle of a drug, and things change. Or we have had some situations where we had some questions imposed and then we had to qualitatively investigate later how that was changing how people might answer things.

So there's multiple inputs to -- there's not just the theoretical underpinning that you come with as a sponsor and how you explore that, but then you have FDA as a stakeholder who is putting questions in as well, and then we have to explore how that changes questions and so things are always sort of moving and it's not just a one-time developmental process.

MODERATOR TOIGO: So it's the practical challenges again, that we keep -- that keep coming up. Okay Dr. Morris.

PANEL MEMBER MORRIS: Well, what if FDA did this? What if they said -- if the

first six months you have to get your interventions in shape and that you wanted an evaluation of the comprehensibility of your interventions, and then, longer-term, will start doing more, you know, behavioral outcome measures, or more REMS-type measures, but to give you the time to develop the interventions, and FDA could, under the law, just give you a six-month evaluation period to do that, and then switch evaluation periods. Is that something that you guys would consider?

MODERATOR TOIGO: Colleagues do you want to comment on --

DIRECTOR MANZO: This is Claudia Manzo. I just want to be clear. So, are you suggesting something more on the order of some sort of comprehension early on and then assessing knowledge or behavior in a subsequent assessment?

PANEL MEMBER MORRIS: Right. Right, so you know, rather than, at this

moment I'm hearing is you can't do your comprehension testing because of the time frames, but FDA could give you time to do the comprehension, so you get your Med Guide to the best it could possibly be, or whatever your other interventions are and then you can, you know, a year later, do your longer term outcome.

DIRECTOR MANZO: I mean it's something we have certainly talked about internally.

MODERATOR TOIGO: Is there -- oh, Dr. Smith.

PANEL MEMBER SMITH: I know we are restricted to talking about surveys but part and parcel of this all has to do with developing risk communication tools that are actually effective and the idea of being able to build in some time to test the comprehension of the tool before -- or the questionnaire, you know, in a reasonable time frame, holds true for those risk

communication tools as well, and the fact that, in most cases, sponsors are asked to create the risk communication tool and send it in with the filing, means that there is just not adequate time to develop a really effective tool and to really put it through the types of psychometric evaluations that are essential in order to make it effective in its risk communication purpose.

So I think that, akin to what Dr. Morris just said, we should be permitted time to develop those tools to a certain level of, you know, validity and reliability. So --

MODERATOR TOIGO: Dr. Willy did you want to comment on --

ASSOCIATE DIRECTOR WILLY: Okay, my question to you is what kind of time frame are we talking about for -- you are talking about two types of studies here. I'd be interested in hearing a little bit more about that.

PANEL MEMBER SMITH: Well you're

putting me on the spot there. I think there's got to be a longer time frame, and I, you know, I'll throw something out there, I'll put a stake in the ground, maybe we should be given, you know, 8 to 10 months to do that.

But the sponsors should be asked to put forward a detailed plan about how they going to develop that tool, what are the methods they are going to use, the types of testing they are going to be doing, and that should be part of the submission, as opposed to the actual final product.

So I'm sure some of the other colleagues around the table may have stronger feelings about that suggested 8 to 10 months, but that's just a thought, but something around that time frame.

MODERATOR TOIGO: Are there -- does anyone else want to comment on -- Dr. Stemhagen.

PANEL MEMBER STEMHAGEN: I'd just

like to clarify, because I am assuming we are talking about after approval, or bridging between pre-approval and post-approval. I don't want it to be sounding like we are now going to be extending the approval time because we have to test all these instruments, because that's going to be a big problem.

So we may be talking about starting, if it's eight months, four months before, but continuing, we are not talking about eight months necessarily before approval.

PANEL MEMBER SMITH: That's correct. That's exactly what I mean. Thank you for that correction.

MODERATOR TOIGO: Dr. Shiffman.

PANEL MEMBER SHIFFMAN: I very much agree with the point that these, both the development of the communication tools or materials and the development of the testing of the test instruments, are serious

developments that take some time, and sponsors need time.

I think there's also a process issue that has been referred to obliquely, which is these are science-based developments that have a real development path, and what I have seen happen and was obliquely referred to, is that the sponsor may do the work and then the agency or individuals at the agency just say no, no, we don't like that question, put in this question.

It would be the equivalent of the agency saying to a sponsor we don't like that molecule. What if you put a bromine group over here? That would look nicer. You know, I think the burden has to be on the sponsor to develop things to be able to show you have made the point that it's not always submitted to the agency. The sponsor should need to show what the process was by which they developed both the communication and the tools, but at the same time, the agency

should be reviewing that process and not ad hoc saying oh no, we want the whole prescribing information in the Med Guide or how about this question I just came up with, if the sponsor has in fact put in the work to develop questions that they have shown work.

So I think both the time line and that process is a matter of recognizing that there's a methodology to doing this, and that the burden to do that needs to be on the sponsors and accountable to FDA.

MODERATOR TOIGO: Dr. Peterson?

PANEL MEMBER PETERSON: Yes, I just want to raise a practical issue and it may be that I am not understanding, because I think it depends on which type of information communication tools we are discussing because the Med Guide is actually part of the label, and it gets developed as part of the labeling discussions, which can come down in a week or two before approval, and honestly, I've seen situations with some drugs where even the

first discussions of the REMS and the REMS goals are really within about a month of the approval.

So I'm not, if we are going to consider this, I think we need to think about the Medication Guide, and then my question would be, if we were doing testing and finding concerns with the Medication Guide or items you would want to change within that tool, I'm not sure what the mechanism is, because it would seem to me that that leads to, in a sense, labeling negotiations which you know, has its own time frame. So I just wanted to raise that at least with relation to the Medication Guide.

MODERATOR TOIGO: Thank you. Dr. Willis.

PANEL MEMBER WILLIS: I would recommend, as long as we are talking about questionnaire design, that one way you think about the development of these items is not strictly within the realm of questionnaire

design for sample surveys, but rather, a kind of different subscience here.

What you are really involved in is the science of test development so that people who work at College Board, whatever, deal with these issues all the time and there's a whole literature there that relates to how one develops tests, how you do calibration, how you determine the difficulty levels of these items so that you know what to do with the eventual numbers that you get out of your investigation.

I heard this morning there's a recommendation down the table a ways that what you should do is try out a test on a group of people who have no particular reason to have any knowledge of the side effects or the risks of a particular medication and see how that group does on your test, in effect.

And if that group comes up and is able to guess, you know, 80 percent correct level, then you know, it's pretty clear that,

you know, 80 percent is not particularly meaningful from your patient group.

So it's that kind of logic and methodology and background science that I think could be really useful for driving the questionnaire development process.

MODERATOR TOIGO: Thank you Dr. Willis. Dr. Shibuya or Dr. Willy, do you have a comment on either? Okay. Dr. Ostrove did you have your hand up? No, okay. Dr. Chowdhury.

PANEL MEMBER CHOWDHURY: I want to go back to that issue, the importance of separating the knowledge, whether it's coming from general practice or it's coming from the level.

I think from a regulatory point of view, from FDA point of view, it's not relevant whether the knowledge is coming from the level or not, as long as the -- it is safe, it can go ahead.

So here's an example. Let's say we

try to separate the effect, separate the knowledge from the practice on from the level. Of course it's very difficult, in some cases it would be difficult to do this, but we do it and we find that from the general practice, their knowledge is 80 percent, and from the level, it's zero percent, still 80 percent, and our endpoint threshold is 75 percent.

In that case, even though the knowledge coming from the level is only zero percent, is FDA going to stop this? Because it's already safe, so there's no point in, you know, separating in knowing this. For that point of view I think that, you know, as long as knowledge is sufficient and risk is minimized, that good enough.

MODERATOR TOIGO: Do you want to comment on that? No, okay. Is there any -- yes. Dr. Fay.

PANEL MEMBER FAY: Yes, not to contradict them too much, but I just want to

caution Dr. Willis that item testing theory and so forth can be overused in this context too. When -- if there are specific goals, specific knowledge goals that you are achieving, this may be contrary to some of the general goals that item testing theory has, in the sample sizes and the methods that are used, not to rule it out, but just be aware that the methods used to develop the SAT might be quite a bit different than what can be applied in this very small context.

PANEL MEMBER WILLIS: I would just respond that even though there may be differences here that are -- I mean, I would call them nuances in some sense, the fundamental point, I think, is that you have got to be aware of this fundamental issue of what's called item difficulty, the need for item calibration.

With any kind of knowledge test, I think it's important to know whether this instrument, with its distractors or

262

true/false or whatever concepts, formats are used, whether this is something that is a fair test or whether in fact it's so easy, say, that it's not really an adequate test of knowledge, that someone could guess their way to an 80 percent level, because that isn't really going to provide you what you want. I mean, something will look like it's at 80 percent, but I just caution you to make sure that that actually represents knowledge rather than guesswork or test-taking ability or some other intrinsic property of the measurement instrument.

PANEL MEMBER FAY: And I'll agree with the emphasis you just gave.

PANEL MEMBER WILLIS: Thank you.

MODERATOR TOIGO: Okay. It looks like we are -- well no. Dr. Willy wants to add some more questions.

ASSOCIATE DIRECTOR WILLY: No, well I would just like to, one more time I am going to ask, if anyone is interested in

speaking to the question that we have about trying to balance developing a survey that really will help us understand the knowledge rate or however you want to describe it, balancing it that with not having too much burden, having very long surveys that people are not willing to complete, having a number of questions around the same particular risk that we are interested in.

So we understand that's a challenging balance and I guess it would be interesting to hear any feedback you might want to share.

MODERATOR TOIGO: Dr. Sleath.

PANEL MEMBER SLEATH: I would just comment that I think that the key thing is you have to get at what are the core messages like what, again, when we are designing studies, you want -- you don't want to power on too many things because the more things you are testing, the more things could look significant by, you know, just statistically

significant by chance.

So I would say you really have to focus on what are the core messages you want to get across, what types of knowledge. I like what he said about the test development but I would think of it more from the educational literature on assessment and Bloom's Taxonomy and are you -- because we are going through curriculum reform, so unfortunately this is on my mind. Do you want to teach basic facts? Do you want to teach where to look for knowledge and how to apply it?

There's -- and I don't have the taxonomy memorized but that might be something good to look at in terms of maybe you have different assessments for different levels of knowledge and I don't think you are getting at just someone said the FDA just wants to know if people can regurgitate facts. I don't think that's what you are after. You are after whether providers can

apply it, whether patients can understand, if they take this with another drug, should they ask somebody or you know, something like that.

So I would look at the educational assessment literature, because it's kind of related to it.

MODERATOR TOIGO: Thank you. Dr. Krotki.

PANEL MEMBER KROTKI: I'll stick my neck out and mention a number, so 10 or 15 minutes. When I -- I worked for Knowledge Networks that does a lot of internet sampling and surveying, and they had a limit of 10 minutes.

So if you -- and I don't know whether everybody would agree with that, but any longer than 15 minutes starts getting into areas of response -- serious response burden and deterioration of quality of response and things like that.

So if you can work with that kind

of parameter, that sort of gives you a framework within which you want to, one, develop your concepts, your goals, your objectives, and two, translate that into questions, be they true/false, or multiple choice, recognizing of course that anything multiple choice or case vignette is going to take longer than a simple true/false question, and yes, the issue response burden, it's the first time it was raised since lunch, and I think that's a very key issue. We need to keep that in mind and keep it -- keep the question as short as possible and as simple as possible.

MODERATOR TOIGO: Dr. Gilsenan.

PANEL MEMBER GILSENAN: Yes, I think we typically try to keep it around 15 minutes, but I think definitely the response burden that you have to balance that with the remuneration. If you will incentivize, if you pay people more for their time, they are more willing to sit down and take the time to

do it, and there is a balance with that as well, and it's not a trite thing to have to balance.

And with, that's mentioned before, with healthcare providers, it's a real challenge to be able to compensate them for a very long time and get their attention for a longer period of time.

And in some of this, you can do some in pre-testing and kind of get a sense of the burden, and I think also it varies on the complexity of the messages that you are trying to get across and the patient populations and, you know, the underlying disease characteristics, as to how, you know, they are able to take a longer survey.

MODERATOR TOIGO: Thank you. Dr. Shiffman and then Dr. Stemhagen.

PANEL MEMBER SHIFFMAN: On the issue of multiple items to measure a single endpoint, I mean, in general in psychometrics, more items are better. If

they are correlated, it leads to a more reliable composite.

But that doesn't necessarily hold for knowledge, and particularly in this context, that is because it's -- but when you score it that you way you are almost making the assumption that knowledge of fact A and knowledge of fact B are interchangeable, that you are going to get the same score if you know A and not B or know B and not A, and I think we need to -- I haven't seen a lot of multi-item composites, but I think one needs to look very closely and ask again from a clinical risk perspective, what is the risk of not knowing a particular fact, and can it really be substituted by knowing some other fact of the same domain, all of which goes to the issue of burden in the sense that I think multi-item scales have to get very close scrutiny as to whether it makes sense to ask all these items about the same thing or you just really -- there really is just one fact

which either they know or they don't, in which case a multi-item scale adds burden without adding value.

MODERATOR TOIGO: Thank you. Dr. Stemhagen, Dr. Raghunathan and Dr. Ostrove.

PANEL MEMBER STEMHAGEN: I'm sure what I'm going to say is pretty obvious, but part of the length o the questionnaire is going to depend on the number of risk messages, so we have some REMS that really are only related to teratogenicity and then you have some that might be teratogenicity and suicidality and whatever, and there might be three or four.

So clearly, to get through all of them, that's longer. We usually find 15 to 20 minutes, and then the demographic questions are at the end, so they are pretty easy and they are not requiring a lot of thought, and no offense, but a lot of the FDA questions about the Med Guides are the next to the end, which again, did you read the Med

Guide, did you, all that, whatever.

And so that sort of goes pretty fast at the end, the key messages about the risks are usually in the beginning, and we don't find very many people who stop the interview. Now again, we -- they are told in the introduction that they are only paid for completed surveys, so that certainly would influence that, but that stop before they finish -- occasionally but it really doesn't happen very often.

MODERATOR TOIGO: Thank you. Dr. Raghunathan.

PANEL MEMBER RAGHUNATHAN: Yes, once you have made a list of all the cold knowledge of awareness that you want to measure, then you can develop a vignette and then follow it by an item, true/false item or you can have a five-point scale item administered.

But also you don't have to measure everybody on all the items, so you can

actually give subsample of items in order to reduce the response burden.

So there are survey designs where people get subsample of items and not that -- but you cover the entire population, but they get like four items, five items, and you can reduce the response burden.

So nobody says that every item has to be administered for every individual that you have sampled.

MODERATOR TOIGO: Thank you. Dr. Ostrove.

PANEL MEMBER OSTROVE: This is a one-off thought that was really brought about by what Dr. Willy said about burden, and it may be completely obvious to everyone.

But one of the burden issues also has to do -- with patients in any case -- also has to do with whether the patient is even the appropriate person to be asking because in many cases, and for many patients and not necessarily elderly ones, they don't

272

even want to know about a lot of this stuff, and they rely on a family member or a friend to read the stuff, to explain to them what the options are, and by asking the patient, you are not even getting the decision-maker in some ways.

So I mean, it's another complicating factor, but like I said, it's just kind of this one-off thought that burden brought around to me. So --

MODERATOR TOIGO: Dr. Holmes.

PANEL MEMBER HOLMES: I'm trying to think of how to best describe this. I think that we have had some experiences where we had single questions that didn't have a lot of psychometric or qualitative work behind it. In particular, I am thinking at the moment of some FDA-imposed questions that we did not understand how folks had answered them until we did qualitative testing the next year on those imposed questions, which brings me back to why I'm very nervous about

single item questions asking a key risk message that we as a sponsor have to report back to FDA on what that rate is, that may be judged against a threshold because if the qualitative work isn't there, if the question has been imposed, and most risks are complicated. And so there needs to be multiple ways of addressing that risk.

So burden is a key factor here, but at the same time, with complicated risks, some of which have imposed questions, some of which has no qualitative work behind it, there is a lot of risk psychometrically and for sponsors as it relates to being so worried about burden that we go to single items.

MODERATOR TOIGO: Dr. Gupta.

PANEL MEMBER GUPTA: Just to add to that, if a questionnaire has been pre-tested, and if you are going to modify any question, add any question, that should really entail another round of pre-testing, because if a

274

questionnaire is no longer valid and reliable, if you have modified any item in any way, shape or form.

MODERATOR TOIGO: Thank you. Dr. Hornbuckle.

PANEL MEMBER HORNBUCKLE: I'd like to just comment on a comment that Dr. Krotki had mentioned as well, which is very practical, that the questionnaires should be very simple, very simple and not as complex.

We have to also think about -- part of the risk message is also, to the patients also, is to understand clinical risk as well as signs and symptoms.

Part of that is in the Med Guide, so we have to be very careful not to have composite questions so that you ask a patient, how many of the risks are you aware of when you take these drugs, and there may be six risks, or there may be seven risks. The more complex you have individuals to select from a composite number of questions

or signs and symptoms, the less susceptible you'll find your response rate. So that's why, as Dr. Krotki had mentioned, I think it's very important to have simple questions and not have a lot of complex, composite scores, where you have to score 6 out of 7 or 8 out of 10 or what have you, but to keep it very, very simple. So keep that in mind, the complexity of having composite questionnaires as well, with the scores.

MODERATOR TOIGO: Thank you. Dr. Willy, did that answer your questions or give you enough to think about? Okay. So we will move on to the last question for this panel, which is question 4. And that's process issues, and we have about a half hour to discuss this question.

And I will read it for the record, unless -- did you have something? No, okay. So the question is: "Please discuss process issues related to these surveys."

Section A, "Given issues of recall,

276

should the lag time between the REMS communication and the survey administration be standardized?"

Part B, "Should pretesting/validation be required to reduce the likelihood of a poorly worded question that was not recognized during survey development?"

Part C, "On average, how long does it take to design, test, recruit participants, conduct, analyze, and report the results of a survey?"

Part D, "Please comment on appropriate incentives for patients and health care providers to complete the surveys."

And so I think we have covered some of these already during the discussion, but now is the opportunity for those that were holding their comments earlier to voice them here, and does someone want to get us started on A, issues with recall?

277

Oh, Dr. LaVange?

PANEL MEMBER LAVANGE: Well, I was going to answer another question.

MODERATOR TOIGO: Okay, well then go right ahead. Go right ahead. You got in there first so pick which one you want to start off with.

PANEL MEMBER LAVANGE: Sort of with the general idea of these, the process and the time line and everything, I just had a reaction to the previous question's discussion. It is obviously important to know the instrument that you are using to assess and we talked about that.

But I do think it's also important to think in advance and get set up how you are going to select your subjects, this whole sampling frame setup and infrastructure, if you are going to do random sample, if you are going to cluster it, how you are going to get your data sources.

And I think that that's as

important as getting the questionnaire right. A good questionnaire is not going to be worth much if the people that answer the questionnaire don't represent the people that you are worried about, the target population.

So I do think that some of this could be done in advance, I know there's a time crunch at the end, and I'm not sure, the sponsors would have to speak to the sequence of events.

But if you are in the survey world you know that you have to set the infrastructure up and identifying sampling frames and things, you know, they don't even necessarily need to be disease-dependent if you are not going for a patient registry, but you are going for a prescribing database or a claims database and so similarly, if you do cluster sampling, you know, there's just lots of infrastructure that could be put into place.

But I just don't -- I do think the

representativeness of the sample, the ability to get a probability sample, all the things that we talked about earlier, should be in this planning stage, not just designing the questionnaires, so I just wanted to make a plug for that.

MODERATOR TOIGO: Thank you. Anyone else want to comment on -- Dr. Stemhagen.

PANEL MEMBER STEMHAGEN: Just sort of a practical response again. It's probably stating the obvious, again, but to that end, we might -- you know, we certainly think about what ths sampling frame is, whether we are going to use pharmacies, or whether it's a product that is buy and build so you have to go to the physician and so on, but you can't go much further than that until the product is on the market because you need to have patients that are using the product. They are the ones who get the Medication Guide and then FDA, this is one question

actually that we have had for FDA because when we started doing some of these, we thought we would be surveying prescribers, and then we got a lot of feedback that it would be anyone who was possibly going to prescribe, so if there was a communication plan it was whoever was informed about the communication plan, and now recently we have gotten feedback again that seems to be 180 degrees reversed and it's only prescribers.

So again, to get prescribers you have to wait until the drug is on the market to prescribe, but you would know where you are going to get them from.

PANEL MEMBER LAVANGE: If I could just, quickly, I mean you also don't market a product until it's approved but you certainly get ready to market the product, and you know, identifying the frame, I mean you can't administer the questionnaire to patients until there are patients, but you can certainly get ready to draw your sample and

get your -- in a sense, the execution, you know, is way down the stream from the planning.

MODERATOR TOIGO: Sure, well you know that you are going to sample through pharmacies and you can get all that set up, or you know you are going to sample by taking a list of physicians who are going to prescribe and you can get that set up.

PANEL MEMBER STEMHAGEN: But you can't get much further until you actually have patients receiving and docs prescribing so there's two steps to it. So I think we try to get as much done ahead of time, of course. It also gets to the survey development as you can have a sense of what hat your questions are going to be, but until the labeling is finalized, you don't know necessarily exactly what the wording is going to be and exactly what things are going to be. So it's hard to get everything set up before all that happens and that's usually

right around the time of approval.

MODERATOR TOIGO: Dr. Raghunathan. Dr. Holmes did you have something? Okay.

PANEL MEMBER RAGHUNATHAN: Going back to this, the point A, that again depends upon whether you want to really measure the long-term retention of the knowledge or you are really interested in knowing how quickly they -- how quickly you want to get into that, so that has a different aspect of knowledge as opposed to sort of a retention of that knowledge.

So you may want to build in sort of some immediate and some lag time, larger lag time as well in order to, if you want to assess how long the person has this knowledge, retains his knowledge, and also it depends upon the drug, right? If it's really a chronic use drug, then the lag time may not be that important because they are constantly using that drug and probably they are getting medical literature from the pharmacy where

they are getting dispensed as well.

So it's a -- the question A is just, it depends upon what do you want to measure? Do you want to measure the long term retention or short term?

MODERATOR TOIGO: Thank you. Okay, Dr. Holmes.

PANEL MEMBER HOLMES: So, just in response to that, Dr. Raghunathan, I recall back to the beginning when Dr. Willis asked are we in a surveillance model or an evaluation model, and I believe Dr. Willy responded that FDA is interested in really the former, the surveillance model.

And so they want -- so you want to know what the knowledge is from all the sources of information, and that there is an interest in just whether there's been exposure to the educational tool.

So the lag time question, then, ceases to be about the knowledge from the intervention, and just how far after exposure

to the educational tool they will retain that memory.

So I guess it gets back to the surveillance model versus the evaluation model in terms of how one answers the question in it right?

So my -- I guess it comes back to am I remembering correctly earlier this morning, are we a surveillance model where we are looking at knowledge from all sources and that the primary question is just exposure to the intervention tool, and therefore the lag time should be close to that intervention, and those questions should be around that.

So it starts to become complicated, right, depending on which model we are using.

MODERATOR TOIGO: Dr. Lee and then Dr. Peterson.

PANEL MEMBER MATTHEW LEE: Well, if I can make this any more complicated, I think one thing we need to keep in mind when we are answering A and B up here and we are thinking

about this, is that REMS in the communication documents and the mechanisms that go with them are fluid. They are not stagnant.

We are updating them all the time. The risk messages change. We get new risk messages based on pharmacovigilance activities and new information that we are finding.

So if we are talking about lag time and how can somebody remember something when we are talking between the risk communication and the survey, well what if that communication changed in the meantime?

There are scenarios, I will give you an example, where a survey protocol was sent in in February and there was a change to the Dear Healthcare Professional letter that came out in March and then the survey was administered in April.

So we get conflicting survey results based on did the healthcare professional read the new letter or did they

read the old letter? You know, so it's you know, one of the things that we are talking about that seems to be important is how long did somebody remember it but is that the most important question? Or the more question is do they know the most up-to-date information? Do they know the newest information and the most relevant information based on all the experience not just the pre-clinical work that went into the risk management plan that was put together prior to product approval.

MODERATOR TOIGO: Well, you were correct. You made it more complicated. Let's see. I don't have anyone with a hand up. Oh, I do -- oh, Dr. Peterson, I'm sorry.

PANEL MEMBER PETERSON: I just wanted to make it even more complicated. The other point to consider is that REMS have a specific set of requirements that we need to do under the REMS, but I know the pharmaceutical companies, we take this very, very seriously, and in addition to what we

are strictly required to do and report on under the REMS, we put in place a whole series of activities to support the REMS program and more importantly, those risk messages. We would typically do that anyway, but even more so.

So sales reps being told all the time to specifically talk about the REMS and the risks, the other methods of communications, promotional materials, some -- we'll develop materials specifically around this to support it. So I think that further complicates that question as well, because this is going on continuously over the life of the product.

MODERATOR TOIGO: Thank you. Dr. Smyth.

PANEL MEMBER SMYTH: I think one of the things that comes out here is that when you try to send those in the real world, it doesn't really work.

And so I think maybe a solution

might be to ask when did you get exposed, not just did you get exposed to this information but when did you get exposed, because then you can say okay, we changed the message so we know who's getting which message, or we didn't change the message and so we can start to look at some of these recall issues, who are the people who seem to know the knowledge and how long ago did they get that information, versus the people who don't know the knowledge and how long ago were they exposed to that information, so I -- it might be worth at least in the early stages to start to approach some of this complexity and try to figure out what's going on before you go to something so extreme as standardizing.

MODERATOR TOIGO: Thank you. Dr. Lee.

PANEL MEMBER SUNGHEE LEE: Sunghee Lee. Again, I am making this a little more complicated. I think somewhere in the document that you sent out, I think I read

that the, I mean, all the survey communities are currently focusing on English-speaking population and I was wondering if there was any plans for either sponsors or FDA to include non- or little English proficiency population, who I would assume that, you know, who are likely to be low on knowledge rate and also probably application of knowledge in their drug use.

MODERATOR TOIGO: Do you want to comment on that?

DIRECTOR MANZO: Hello, that does make it more complicated. We don't, I mean, we have often had companies ask us and wanted to submit materials in Spanish for instance and we don't have the expertise to be able to review that to determine whether it was adequately translated. So we have essentially left that up to the company to sort of state that that information was indeed appropriately translated, and likewise I don't know that we have -- we certainly

haven't asked for specific populations. We have asked again that the surveys be conducted in a representative sample, if that happens to be individuals that are -- don't have English as a first language then I suppose it would include those individuals.

MODERATOR TOIGO: Dr. Gilsenan.

PANEL MEMBER GILSENAN: If I could elaborate on a topic that we have made very complicated in the lag time. It was mentioned the disease states, is it a chronic disease state, an acute, or intermittent even, something that was taken on an as-needed basis.

You know, we looked at across several surveys and knowledge seemed to be higher in patients when it was a chronic condition and a very severe risk as opposed to an acute, like for an antibiotic that is taken very short term, that -- and the risk might not be as -- perceived as severe, and there's a -- how recent did they, if it was

as-needed medication, how recent was it since they last picked up the prescription?

And the other component too, is, and I don't think we have mentioned this so far, and it's really more with the chronic, is it a new user, first-time user, or is it someone who has been using the product for 6 to 12 months or longer?

And we have also found there's differential response to specific questions, and it varies depending on what the question is based on the duration of use experience by the patient.

MODERATOR TOIGO: Thank you. Dr. Cantor?

PANEL MEMBER CANTOR: Yes, I was just going to comment, I think perhaps going along with a couple of the previous comments, that it would be sufficient, A, to ask -- to note in the analysis when people actually responded to sort of get an idea of what the recall period was, but it seems to me that

the critical thing is you want a minimum recall period so that they have had enough time to get the documentation but the amount of time that elapses seems to be more of a substantive concern. So you want people to retain it as they are taking the drug. So it's not a memory issue as much as, is this being retained over a certain period of time, and so there, I would set a minimum period for when you would actually send out the surveys and then note when they actually responded so you can get an idea of long a period that really is.

MODERATOR TOIGO: Thank you. Anyone else on any parts to this question? I think, so Dr. Willy, we had comments before on incentives. We have heard about timing. Is there any one of these that you want to specifically get some more input on?

ASSOCIATE DIRECTOR WILLY: Well, I think we have covered most of these questions and people want to expand on -- we have had

some people share with us how long it takes to come up with the design and testing the tool and so forth, if there's any additional information people want to share in that area. Other than that, I think we have heard really quite a bit already in this area.

MODERATOR TOIGO: Okay. Before we -- oh, Dr. Shiffman.

PANEL MEMBER SHIFFMAN: Just quickly, I think this has been said, but since the question is posed as a yes/no question for B, I would say the answer is a slam dunk yes, that really, I'll put it very strongly, no questionnaire, no instrument should go out without having been pre-tested any more than you would approve the drug with no testing, and that time has to be allowed for that.

With regard to testing process and standards, again, another part of FDA has issued guidance on patient-reported outcomes. My view would be that the principles in that

guidance apply, but the details probably not, because that's not oriented toward assessment of knowledge about a particular drug, and it outlines frankly some very ambitious test development processes that in our experience take a year and a half and one to two million dollars, so it's a different kind of ball game.

Nevertheless the basic principles of qualitative testing and development, it has guidance about validating translations which we just talked about, I think the basic psychometric principles in that guidance apply here even if some of the details are more oriented toward, if you will, more complex and subjective constructs than the knowledge we are assessing here. But I think pre-testing is just absolutely essential.

MODERATOR TOIGO: Okay. We have -- we are a little before time. So two things. We have some panel members from panel 1 who are not participating in panel 2, and so I

think if there are any, now would be your chance to provide us with any comments that we -- you want to make sure we hear, and this would be a good opportunity to do that.

Okay, Dr. Fay.

PANEL MEMBER FAY: I'll just point out an ambiguity even not wishing to disagree with the last comment, but just differences in culture.

In survey research, pre-test may sometimes mean sample sizes of 300. So I'm not sure whether you had that -- as opposed to development might be little tests of 30 or 10, often done in cognitive testing, is -- the cognitive testing may be separate from the pre-testing, which is done on a larger scale. So I'm not quite so sure the answer to B should be yes do pre-test at 300.

PANEL MEMBER SHIFFMAN: No, you're right. In a context where a sample of 300 would be pretty good for the final test, you wouldn't want to dictate that for the pre-

test. So I had in mind the smaller scale things which unfortunately still sometimes get pushed out by the aggressive time line.

MODERATOR TOIGO: Okay then I am going to use this opportunity to just, we are not going to comment on it -- oh, I'm sorry. Well, before I do that, Dr. LaVange has a question or a comment.

PANEL MEMBER LAVANGE: Right, because I'm not around for the second --

MODERATOR TOIGO: You're not back, okay.

PANEL MEMBER LAVANGE: But that reminded me of, just following up on Dr. Fay's and the pre-test sometimes also refers to not just the questions, but the operation of the survey.

So the other, getting back to this testing of the infrastructure and so forth.

But we talked this morning about representativeness of the sample and then I was just, I wasn't sure about the answer to

the Spanish and other non-English language questionnaires, because if the population that is being prescribed does include Spanish speakers or Spanish as a first language or non-English speakers of any sort, then you are not obviously going to have a representative sample if you exclude that population. That would just be like excluding people that don't understand what you are trying to assess if they understand.

So, but also, having had to do this in an academic environment, and not having the expertise to review the translations as you mentioned, there are certification -- there are groups that can certify a translation. So I think there are some things, if the FDA were interested, that they could require in terms of a translation that was certified by an outside party or something like that, so that everybody developing these questionnaires doesn't have to have the expertise to make sure, you know,

to review it, to make sure that it's okay.

But that does get to the representativeness. So you wouldn't want to exclude, I don't think, whole segments of your target population, right?

MODERATOR TOIGO: Dr. Stemhagen.

PANEL MEMBER STEMHAGEN: I think sometimes, to that point, that the sponsor takes the position that the Medication Guides, if we are talking about patients, are only available in English. And so there aren't the opportunities for Spanish education and using that method, and questionnaires get developed primarily based on the Medication Guide, that's been sort of the reasoning about having it in English.

It does lead to this bias, as we are talking about, but we are not providing the materials in another language.

MODERATOR TOIGO: Dr. Ellenberg, were you going to -- did you want to say something?

PANEL MEMBER ELLENBERG: Yes. Thank you. Jonas Ellenberg. What? The Med Guides don't come out in other than English?

PANEL MEMBER STEMHAGEN: Well, I'm not a sponsor. I work with a lot of sponsors. I think there are some that do but primarily they are in English.

DIRECTOR MANZO: I just want to mention that we don't preclude companies from putting information out in different languages.

PANEL MEMBER MORRIS: It's required to be in English. Companies can do anything they want. But the legal requirement for labeling in the U.S. is English.

PANEL MEMBER ELLENBERG: But in pursuing whether this works, you are assuming that all Hispanic or anyone else speaking Spanish will be fluent in the English language, enough to be able to read this document. Certainly the physicians but not the users of the medication. That's the

assumption and that's --

PANEL MEMBER STEMHAGEN: The patient or their designated care giver, whomever, yes. I mean that's the assumption if you have your survey being done in English, that's correct.

PANEL MEMBER MORRIS: I mean there's a lot of people excluded beyond Spanish. There's people who are blind. There's people who can't read English for whatever reason, they are illiterate, I mean they are not in the target group either. They can't read a Med Guide. They have to have either a caretaker give it to them or someone explain it to them or have some other intervention, but there's limits to what a Medication Guide can do.

MODERATOR TOIGO: Dr. Lemeshow.

PANEL MEMBER LEMESHOW: So I just have a question. Is it possible, if we think about a drug that is considered potentially dangerous enough to require REMS, is it

possible to think about creating a sampling frame on the run basically, so that when a patient picks up the drug at the pharmacy, just a simple, you know, their name is written down or something so we have some sort of a registry? Is that, because if we had a registry, then a lot of the issues that we talked about would be easier to deal with. Without a registry then there's a lot of complications. I don't know the answer. I am just asking if that's even possible.

MODERATOR TOIGO: There are registries for some of these drugs, but do you want to comment further on that?

DIRECTOR MANZO: We haven't been requiring registries unless there was some type of safe use component, so not sort of registries of every patient that would be taking it for the sole purpose of assessing whether the REMS is working. That might be a little bit intrusive.

MODERATOR TOIGO: Well I think, to,

I think what you are saying is that the REMS, there is a goal for the REMS and if in order to meet the goal, a registry is a requirement, then that becomes part of the REMS. So we are looking at burden, balancing access and burden, and so while you may be correct, we are dealing with a bunch of things when we are trying to meet the goal of REMS. And we can only -- we only want to require what's necessary in order to meet the goal.

PANEL MEMBER LEMESHOW: So, my whole point here is that it may be not possible to really get a representative sample without something like that. So if you are going to require the REMS, what about the next step or sort of a preliminary step to make evaluating the REMS -- evaluating the instrument possible.

MODERATOR TOIGO: I think part of the -- I'll let my colleagues speak as well, but part of the purpose of this workshop is

to, now that we have experience with REMS, we have four years of experience in implementing them, and many of the first ones were Med Guide REMS, so we were evaluating Med Guides and all of those have gone away from -- if it's only a Med Guide-related REMS, that will no longer be a REMS, so we are not doing assessments of Med Guide-only REMS.

And so now, it's a point in time for us to step back and say: what have we done, what's working, what's not, what are the challenges that we maybe can address, and then acknowledge that some of these challenges are just going to remain challenges for us to work with. So I think my colleagues may be able to say it better than me, but I think that's as best as I can get, from where we're at.

ASSOCIATE DIRECTOR WILLY: This is Mary Willy. The only thing I would add is that since we are moving in a new direction in a sense, so the Med Guide-only REMS are

moving away. The REMS that we do have are -- generally have elements to assure safe use, and we -- that tends to give us more information and maybe we will have more opportunities with those types of REMS to collect data that we have had more challenges with when -- with the Med Guide only. And then this afternoon we are going to be talking about other ways of maybe evaluating some of these REMS, using different kinds of options that we haven't used so far.

MODERATOR TOIGO: Right. The last session is to look at alternatives to surveys. We needed to spend a significant portion of the meeting on questionnaire design and all of the things related to that, but the last panel is to address: well, what are our other options to look at evaluating whether or not REMS are meeting its goal -- their goals.

So I think that's probably a good place for us to stop, and I just want to read

the one question that I have already received into the record. We are not going to talk about it, but just to have it acknowledged.

The question comes from Monica Groat from Sidley Austin. "As has been discussed, it is common for REMS to require manufacturers to assess the effectiveness of mandated communication plans. Has FDA established any standards for determining whether a communication plan has been successful?"

So I think we have got that comment. Unless one of my panel members has a quick response, I think we have read it to the record, and acknowledged it and if other questions from the audience, please drop them off with Julia.

So we will see you at 2:15. Yes, 2 to 2:15 is this break.

(Whereupon, the proceedings in the foregoing matter went off the record at 1:57 p.m. and resumed at 2:16 p.m.)

MODERATOR TOIGO: Okay, we are going to get started with panel 2, and Dr. Robert Shibuya, who is the deputy director from DRisk, is going to briefly summarize the discussion from panel 1, and then he is also going to describe the focus of the second panel, and while he comes up here and gets ready, we are going to go around and have the panel members introduce themselves, because we have some new panel members for number 2, and those of you who participated in panel 1, if you wouldn't mind just reintroducing yourselves again.

So we'll start down at the end of the table with Dr. Wolf.

PANEL MEMBER WOLF: Mike Wolf, Northwestern University.

PANEL MEMBER DAVIS: Terry Davis, Louisiana State University, Health Sciences Center, Shreveport, Louisiana.

PANEL MEMBER DAY: Ruth Day, Medical Cognition Lab at Duke University.

307

PANEL MEMBER EGGERS: Sara Eggers, Office of Planning and Analysis within CDER, here at FDA.

PANEL MEMBER GILSENAN: Alicia Gilsenan, Senior Director, Epidemiology, at RTI Health Solutions.

PANEL MEMBER GUPTA: Sanjay Gupta, Daiichi Sankyo.

PANEL MEMBER HOLMES: William Holmes, AstraZeneca Pharmaceuticals.

PANEL MEMBER HORNBUCKLE: Ken Hornbuckle, Eli Lilly.

PANEL MEMBER KROTKI: Karol Krotki, Survey Statistician, RTI.

PANEL MEMBER MATTHEW LEE: Matthew Lee, Lundbeck.

DR. IZEM: Rima Izem, the Office of Biostatistics, at FDA.

MR. GORDON: Brian Gordon, Division of Risk Management at FDA.

ASSOCIATE DIRECTOR WILLY: Mary Willy, DRisk, FDA.

308

PANEL MEMBER RAGHUNATHAN: Terry Toigo, Associate Director, Drug Safety Operations, CDER.

DR. AUTH: Doris Auth at the Division of Risk Management.

PANEL MEMBER STEMHAGEN: Annette Stemhagen, United BioSource Corporation.

PANEL MEMBER SMITH: Meredith Smith, Abbott Laboratories.

PANEL MEMBER SHIFFMAN: I'm still Saul Shiffman, University of Pittsburgh and Pinney Associates.

PANEL MEMBER MORATTO: Elaine Moratto from the Colorado School of Public Health.

PANEL MEMBER PETERSON: Janet Peterson, Janssen Pharmaceuticals.

PANEL MEMBER OSTROVE: Nancy Ostrove, independent.

PANEL MEMBER MORRIS: Lou Morris, LAMA.

PANEL MEMBER LEFKOWITZ: Doris

309

Lefkowitz, Agency for Healthcare Research and Quality.

PANEL MEMBER LEMESHOW: Stan Lemeshow, Ohio State University.

MODERATOR TOIGO: Okay, Rob, do you want to get started?

DEPUTY DIRECTOR SHIBUYA: Okay, good afternoon. I wanted to start by thanking the first panel for a very interesting and wide-ranging conversation this morning.

I was -- I have been asked to summarize the sort of consensus or what the major points were from the panel 1 discussion for the benefit of anybody who just arrived.

And as a non-social scientist and non-statistician, I hope I have captured the key points. I have done my best.

We had four topics, basic topics for discussion this morning. The first was endpoints. There was not a whole lot of consensus. I thought that there were a

couple of points where there was consensus. It seemed that people thought that the messages to be tested needed to be risk-based. It felt like if we could identify -- we were told that we should identify the key risks and if we do that, it should be fairly straightforward to set a threshold and then after following that, the sample size would just be a calculation.

There was a lot of emphasis on making sure that the educational materials were high quality, there was construct and content validity and they were well developed.

At the -- during the first session there was also some concern raised about the non-response bias, which occurred in the subsequent sessions as well.

FDA was asked to consider a couple of things. We were asked to consider sort of what our goals were, what the research questions would be.

Do we want to establish sort of causality, whether or not the REMS is causing patients and prescribers to have this knowledge or not, and then there was also a question of whether or not we wanted to look really as a snapshot -- or I think it was called the surveillance model versus the educational model -- which of course we'll take back and vet internally.

And then the other issue that came up, that we need to consider in the endpoint discussion, was whether or not these should be conceptually open or closed books, and we will take those comments back as well.

The second area was -- sorry, not the second panel, but the second group of questions -- pertained to recruitment and sample size.

We were told that theoretically this is easy to do. You just define your sampling frame, you pull a -- I think a random sample -- you get your estimate of

312

variance, and then you do the sample size calculation.

But then amongst the panel members, there were issues of feasibility or practicality that were brought up. Again there were questions about the non-response bias or concerns that were brought up, issues of feasibility, because we can't really compel -- or sponsors can't compel -- invitees to actually participate, it's difficult to get a random sample.

Industry pointed out that the target populations can be difficult to define, because the information sources only usually give very vague demographic information, usually limited to age and gender.

The other issue that came out is issues of burnout, and whether or not there could be enough remuneration in order to keep participants interested in doing the studies.

And then other thing that came up

313

was concerns about time frames and the time to recruit these participants.

The next group of questions pertained to question design. There was not consensus there. Some panel members had a lot of support for the open-ended and clinical vignette-type questions, but acknowledged that there were drawbacks with regard to the resources and the interpretation, getting the burden, getting people to participate. So there was also support for very simple, true/false type questions, but of course it was emphasized that it's very important to properly test these things. But there was an argument made that it needs to be appropriate, the question type needs to be appropriate to the population being surveyed.

Again -- actually the question, we were questioned again -- are we interested in more the regurgitation model, or the retention-of-knowledge model, so again we

will take that back.

And then during the question design discussion, there was a discussion of the -- I think it's called response burden -- and we asked the panel for advice about balancing the response burden versus getting a lot more information.

And then the last group of questions pertained to process. With regard to lag time, the same question was raised, again, because obviously a long lag time would imply that we are interested in retention of knowledge versus sort of a snapshot or really comprehension of the materials.

There was a strong sentiment that pre-testing of these materials is essential. Anything else in here -- oh, there were also issues raised about the time line, and it was clear that in order to produce really quality materials, it takes time. There was mention of a model where perhaps we could start,

315

maybe the assessments could be phased with simpler things like finding out whether or not the materials are comprehensible, and then later looking at more complicated outcomes such as behavior.

And then last there was a concern raised about the fact that the materials are basically only in English, and whether or not that could cause problems with bias, in terms of the interpretability of the results.

So with that, I was going to go on and introduce the panel 2.

Good afternoon. In this presentation, I will provide an introduction to panel 2, entitled "Alternatives to surveys to assess knowledge and the use of surveys to assess behavior, access, and burden."

Next slide please. Oh, I can do that. I'm going to cover the following topics: first, I'll charge panel 2 with discussing two issues; then I'll explain the impetus fo this panel; following that, I'll

316

answer clarifying questions. Prior to projecting the questions, either Terry or I -- although I think it's going to be me -- will read you language from the actual law that authorizes FDAAA to emphasize a point.

The goal of panel 2 is, as you can read there, it's an exploratory discussion of the two topics below. For those of you who were here this morning, the goal is substantially different.

We are here for you to explore these topics, and we are not seeking specific ways to improve methods. Your charge is to discuss these two topics, and I'll discuss them a little bit more in a second.

So you might ask yourself: why did we include this panel? If you were here this morning, we had more than three years' experience reviewing surveys done to assess REMS. Several substantive issues, concerns and questions have been raised regarding these surveys.

317

These concerns beg the question of what else could be done to assess whether patients and HCPs have the desired knowledge. With regard to the second broad discussion topic, that I'll call behavior access and burden, there are two reasons we'd like to hear your views.

First, while knowledge is important, behavior is really where the rubber meets the road. Up to now, REMS surveys have rarely included questions involving behavior. We are now interested in how surveys could be useful to assess behavior.

With regard to access and burden, addressing these two issues is a concept that is contained in the law that authorizes REMS. We have also heard substantial feedback from stakeholders in this area.

I'd like to expand on the subject of alternatives to knowledge surveys. To this point in time we have accepted surveys

318

to assess knowledge. This policy is strongly supported in the social science community, as we heard this morning.

In addition, during the July 22 public -- sorry, July 2010 -- public meeting on REMS that we conducted, various speakers described the appropriateness of surveys to assess REMS.

Here I summarize some of the key concepts conveyed at that meeting that you can read. I note that the speakers did note that survey data do not necessarily inform behavior.

So while surveys are clearly an acceptable and established methodology to assess knowledge in populations, as I covered in my introduction to panel 1, we have observed issues in the areas of sample size and the other areas, the other issues enumerated here, that have led to questions about the validity of the survey data submitted, and what else could be done to

assess knowledge.

We are very interested to hear your thoughts regarding possible alternatives to surveys.

I have already covered why we are interested in hearing from you about the use of surveys to assess behavior, so there's no need for further introduction to that part of your charge.

However, we are asking you to discuss how surveys may be used to assess the effects of REMS on reducing patient access to the drug, and the burdens of REMS to the healthcare delivery system.

The law does not address the burden of conducting the survey or the assessment itself. We are not asking you to comment on that in this session.

So, briefly, the law introduces the following concepts that I have here, and this is specific with REMS with elements to assure safe use, but it's the basic concept that's

throughout the law.

So REMS really are not to be unduly burdensome on patient access to the drug, and they are not -- they are supposed to be designed to minimize the burden on the healthcare delivery system.

From the law that I just read, you can see that Congress anticipated potential negative effects of REMS, and asked us to consider them.

We have also heard from stakeholders in this area. Again I refer to the July 2010 public meeting where we heard -- I'll say complaints -- about issues of burden and access. I will also refer you to a July 27, 2011 American Society of Clinical Oncology
meeting that was on REMS, where we also heard about issues of burdens to their healthcare delivery system.

And last, I will -- I put down some key points from -- that we heard from

panelists at our advisory committee that was held at the end of last year, where the iPLEDGE or the isotretinoin REMS was discussed.

And I'll go ahead and read sort of what the main conclusions in this area were. We were told that better data are needed to assess the actual burden to patients. The panel noted that no data had been collected to really address the question of burden. We were told that there is insufficient information to address if prescribers are opting out or if patients are in some way sort of being denied the drug, and we were also asked to collect information on other liabilities brought on by REMS, such as cost.

I guess I can take questions, clarifying questions now. Yes.

PANEL MEMBER DAVIS: Terry Davis. I have one question from this morning. I am not clear how patients were surveyed. Were these paper and pencil or electronic, or were

there structured interviews given over the phone?

DEPUTY DIRECTOR SHIBUYA: Brian?

MR. GORDON: From our experience we have seen all different types of surveys, conducted over the internet, mail surveys, telephone surveys. So they have used all sorts of methods to conduct them.

PANEL MEMBER DAVIS: Just pointing out that many of those are literacy-dependent.

DEPUTY DIRECTOR SHIBUYA: Okay. If there aren't any more clarifying questions, I wanted to just have a prelude to question 5, and then I'll turn it back over to Terry.

So as a prelude to question 5, we wanted to provide you with language that describes what a REMS assessment should consist of, from the law.

And you can read it there. It's notable for saying almost nothing, I think. As you can see from the relevant excerpts,

Congress was extremely vague when describing how REMS should be addressed, and the point of showing you this is: we want you to be similarly broad-ranging in your discussion this afternoon. Thank you.

MODERATOR TOIGO: Thank you, Rob. Let's see. So Cheryl, are we ready then for the open public session?

Okay. So we are going to move to that session, and as I mentioned in the morning, this is a part of the process that we consider very important.

We -- the insights and the comments provided by our speakers in the open public session of the meeting provide useful input for us to consider and for the panel to consider as they deliberate on the questions.

And to facilitate a smooth process, if you weren't here this morning, you are going to have -- each of our speakers this afternoon has three minutes. You will -- the timer will go on and when it turns yellow you

324

will have 30 seconds left -- oh, when it turns yellow you will have one minute. That's changed.

And then when it turns red, it's not going to cut you off, but if you go a whole lot longer I'll probably cut you off. So we're going to try and keep it to three minutes, but we are going to try and do that nicely.

And when -- after you are done speaking, I will ask the panel if they have any clarifying questions, and then if not, we will move to the next speaker.

So our first speaker is Carl Kraus from Medscape. Welcome, Carl.

DR. KRAUS: Thank you. Good afternoon. I am Carl Kraus, Vice President of Medical Affairs, Medcape WebMD.

We serve as an online educational platform for healthcare consumers and providers. Thank you very much to the organizers for allowing me to speak today,

325

and to the attendees for listening.

I am going to provide a high-level overview of data elements available for continuing medical education providers, that can impact REMS assessments.

In an ideal world, after NDA approval, a new drug would be used as detailed in the product label, for the targeted patient, at the recommended dose and recommended duration.

Such ideal safe use is difficult to achieve at the time of marketing authorization or thereafter. REMS is a regulatory tool to ensure that safe use conditions are maintained and that drug access is not limited.

Educational elements of REMS, including Medication Guides, communication plans or training programs, are the primary educational interventions for this purpose.

For the ACCME, CME, a requirement for most states' maintenance of licensure,

326

should be designed to change knowledge and overlapping goal of REMS, as echoed in the 2009 draft REMS guidance.

Who should be educating physicians on the safe use of drugs has been the focus of discussion for well over 50 years. Currently CME providers are being leveraged in this capacity, as denoted by the FDA in the CME blueprint for the class-wide, long-acting opioid REMS.

Although REMS educational requirements do not historically have a CME component, CME activities, delivered electronically, permit capture of data directly related to the respective risk content.

Such data include healthcare professional engagement, knowledge acquisition, knowledge application through clinical vignettes, as well as evaluation of varying learning formats.

There are useful variables

327

collected by CME providers when considering REMS success. FDA-approved content that will be transformed into CME activities independently need to be evaluated for fidelity, because CME providers are not charged with meeting REMS goals. Manufacturers are.

Healthcare professionals that are evaluated before and after an activity permit assessment of knowledge acquisition and application and content question quality.

As an example, on this slide, the third row down, a non-Gaussian distribution of responses mainly imply poor question or content development.

Categorizing participants based on linkage of learning strata permit categorization of individuals into folks who have improved their knowledge, reinforced their knowledge, or had no effect, enabling a re-targeting strategy as well.

There are many features unique to

328

CME data assets that can facilitate improvement of a safe use of marketed drugs and impact the assessment of REMS.

I just want to highlight five. One, engagement, is invaluable so that messaging can be enhanced and targeted.

Two, context regarding participants such as geography, specialty, or care patterns can impact generalizability.

Three, collecting a large sample rapidly, often in excess of 20,000 healthcare professionals, provides a broader landscape for evaluation in characterizing a sample frame.

Four, monitoring outcomes in essentially real time can alert stakeholders to areas that require focus.

And five, testing -- prior to implementation -- testing varying teaching formats and questions to optimize safe use communication and education can be effectively assessed, including use of

feedback, clinical cases and knowledge retention, through targeted re-engagement. Thank you.

MODERATOR TOIGO: Thank you, Carl. Our next speaker is Michael Sprafka from Amgen.

DR. SPRAFKA: Good afternoon. I want to begin by commending the FDA on this workshop and its willingness to discuss how REMS are being designed and implemented.

Amgen has accumulated significant experience with REMS since their inception in 2008 and, as a result, can offer the FDA panel insights into the development, implementation, and assessment of these programs

Our experience with REMS includes elements to assure safe use, and risk communications, with follow-on patient and physician surveys to assess understanding of product risk.

Having the necessary tools and

clear, measurable objectives are essential to the success of REMS. These strategies must be standardized and integrated into the healthcare delivery system to make them work; otherwise they may be viewed as burdensome, disconnected and distract from discussions between patients and their doctors about appropriate choice and use of medicines.

REMS to date have focused in large measure on the transmission of information. While knowledge acquisition by the patient and provider are critically important, knowledge alone is not necessarily sufficient to modify the behaviors that can lead to mitigation of a risk.

Thus, we would encourage the development and use of metrics that address appropriate use of the product, based upon its labeled indication, contraindications and monitoring requirements.

Education materials, when they are used, need to provide information on both

benefits and risks, and must be presented in a manner that is easy to read and understand if it is to be of value in decision-making.

One way to increase the acceptability of REMS programs is to ensure the risk mitigation strategies fit into physician practice and the way healthcare is delivered.

It is why we strongly believe that patients and providers and -- where appropriate, payers -- need to be part of the design and implementation of these efforts.

By focusing on the identifiable risk that can be mitigated, by actively involving other healthcare stakeholders in the development and implementation, by ensuring consistency in the application of educational tools, and implementation of elements to assure safe use, REMS programs that have both meaningful metrics and fit nicely into the way healthcare is practiced and delivered, can be achieved.

Amgen looks forward to participating in these efforts. Thank you.

MODERATOR TOIGO: Thank you, Dr. Sprafka. Any questions from our panelists? Okay. Our next speaker is Dr. Adams from Covance.

Welcome back, Dr. Adams.

DR. ADAMS: Thank you. Again I want to thank the FDA and the panel. I am going to address assessing patient behavior through surveys, and I believe that carefully constructed surveys can be used to address, at least in part, some of the issues involving compliance -- for example, compliance with medications, proper patient selection, and the impact of access of certain REMS programs on patients and on providers.

For example, we routinely use an algorithm in one of our ongoing studies to assess compliance with contraception requirements in a particular program, and we

do that within the context of ongoing surveys where we look at issues such as the age, menopausal status, the use of primary and secondary contraception, et cetera.

The same type of approach can be used with surveys to look at whether or not patients are being properly selected. You could do that with physicians and with patients.

What you could do is ask about age, concomitant medications, other conditions and perhaps recent surgeries, and then create various risk categories, so patients at no risk, moderate risk or high risk categories.

With regard to access, at the meeting last December, the FDA came up with ten questions regarding whether access is reduced or not, and I think a couple of those you could actually address in part with surveys.

One is: are patients willing to put up with the requirements of a survey? You

334

can clearly ask a number of behavioral questions about burden, about time spent doing certain aspects of it, about multiple visits to the doctor, et cetera, to assess the burden.

Another one of the questions, ten questions that were raised by the FDA, is publicity, either positive or negative. And several years ago, in response to a lot of negative publicity on a drug, we asked a series of questions of physicians and a series of questions of patients, and we asked the patients about their use of the drug. We asked them how their physicians seemed to feel about prescribing the drug to them. And then we asked physicians about whether they were aware of the publicity and how they were treating their patients.

And basically what we found was the patients said the physicians seemed more reluctant to prescribe the drug, and the physicians were telling us they were doing

more referrals if they thought the patient needed chronic opioid therapy.

So the point is, you can use this type of survey to address parts of these ten questions that were raised by the FDA in the December meeting.

MODERATOR TOIGO: Thank you, Dr. Adams. Any questions for Dr. Adams? Okay, our next speaker then is Michael Cangelosi from the Center for the Evaluation of Value and Risk In Health.

MR. CANGELOSI: Good afternoon and thank you for your time. My name is Michael Cangelosi and I am here today to discuss research that identified appropriate frameworks for opioid REMS risk communication programs.

This research was conducted with colleagues at the Center for the Evaluation of Value and Risk in Health, Tufts Medical Center in Boston, and was funded by an unrestricted grant from Mallinckrodt

336

Incorporated, a Covidien company located in Hazelwood, Missouri. Mallinckrodt is a manufacturer of prescription pain meds and has paid my travel expenses to attend this workshop.

And we initiated this research by asking the question: given that REMS should mitigate risk in part through communication, how can we best influence behavior through such communication programs?

We sought to identify types of communication programs that discourage risky behavior, which is the ultimate purpose of opioid REMS.

In sum, we identified three key points. First, different communication frameworks can more effectively influence behavior.

Second, REMS programs should be evaluated based upon their impact on such behavior, and not just upon their impact upon patient knowledge.

337

And third, properly designed, communication programs can influence behavior. That is important because, by influencing behavior, risk communications can help mitigate risks.

As a brief introduction, communication of frameworks determined program design. For example, the social learning theory suggests patients would learn and adapt a behavior by watching peers. In contrast, this transtheoretical model suggests there is an iterative process of behavioral adoption.

The key is to use the right framework for the job. In the case of opioid REMS, we want a framework that is effective at discouraging potentially addictive behaviors.

The review process, to identify which framework is best at the job, we searched the HealthComm Key database, a database of peer-reviewed health

communication programs and searched for programs that discouraged any kind of risky behavior: for example, smoking cessation.

And for each framework, we recorded the change in the proportion of subjects successfully influenced for that study; then across those studies, we took the median of those proportions for each framework.

Our review identified 263 abstracts, 126 full-text articles, of which 31 met inclusion criteria, and we found that the most effective communication framework to discourage potentially addictive behavior was that stages-of-change, transtheoretical model.

In sum, properly designed and assessed, campaigns to increase awareness of risk can influence behavior, and in order to best inform future REMS programs, REMS assessments should consider program design, knowledge transferred, and behaviors adopted. Thank you for your time.

339

MODERATOR TOIGO: Thank you, Mr. Cangelosi. Questions for our speaker? Okay, thank you. Then our next speaker is Dr. James Frame from the David Lee Cancer Center, Charleston Area Medical Center, and also representing ASCO.

DR. FRAME: Good afternoon. My name is Dr. James Frame. I am a practicing, community-based oncologist representing the American Society of Clinical Oncology, or ASCO.

With over 30,000 members, ASCO is the world's leading oncology professional organization, representing physicians who care for people with cancer, with a history of leadership and improving patient care.

On July 27, 2011, ASCO sponsored a REMS workshop to address oncology community stakeholders' concerns regarding implementation of REMS programs.

Today, I will present highlights from this workshop that addressed REMS

340

challenges and assessments, as well as strategies, present and future, for REMS assessments.

As shown on this slide, workshop stakeholders included professional societies, advocacy organizations, the Brookings Institution, FDA, the Institute of Medicine, and industry.

Outlined on this slide are a number of identified REMS challenges. Many of these challenges were not identified through current REMS assessments.

From the patient perspective, REMS educational materials often exceed general patient population medical literacy and do not include information on potential benefits or alternatives therapies, focusing only on risks.

From the provider perspective, the devotion of significant time to REMS-imposed administrative duties negatively impacts time available to engage in direct patient care.

341

A number of challenges related to REMS assessment programs were identified. In addition to concerns regarding the unavailability of REMS assessments for independent review, there are concerns that measures of patient comprehension and decision-making may be confounded by disparities in literacy levels and the exclusion of benefit alternative therapy language in the patient educational materials.

Further, patient and provider compliance with the logistical aspects of REMS program implementation as a surrogate for drug safety addresses only a portion of the drug safety continuum.

While there are a number of alternatives to current REMS assessment strategies, there are transformative approaches that ASCO is uniquely positioned to address.

For example, a rapid learning

342

system now in its pilot phase will leverage today's technological advances to better assess and show real-time care data assembled in a central repository that links oncologists, researchers and patients in a continuous cycle of learning.

This will permit learning from each patient's experience, measuring quality and performance and providing decision support in real time.

In the context of today's workshop, the rapid learning system model has the potential to capture drug safety signals in a fashion that may transform current REMS assessment methodologies.

Healthcare provider organizations are uniquely poised to collaborate with FDA in this endeavor. We will submit more robust comments to the docket. Thank you for your time.

MODERATOR TOIGO: Thank you, Dr. Frame. Questions for Dr. Frame? Okay, then

343

our last speaker is from Biogen Idec, Dr. Lynda Cristiano.

DR. CRISTIANO: Thank you for the opportunity to speak with you today. My name is Lynda Cristiano. I am medical director in safety and benefit risk management at Biogen Idec. This is my disclosure slide.

I am going to speak to you today about our experience with TYSABRI and the treatment of relapsing multiple sclerosis and the TOUCH prescribing program, a REMS with elements to assure safe use that was implemented in the United States since 2006.

We have heard a lot about the objectives of REMS assessment, and to date we agree that the primary focus has been on evaluating the knowledge of healthcare providers and patients, rather than on attitudes or behaviors.

Between 2006 and June 2011 we conducted surveys of MS prescribers and infusion sites enrolled in the TOUCH program

every six months for a total of nine surveys, and in recent years this has been done in collaboration with RTI Health Solutions.

This slide shows some examples of the knowledge questions on our TYSABRI prescriber surveys, primarily focused on assessment of knowledge of risk and risk factors.

This slide graphically depicts the results of our surveys, assessing MS prescriber knowledge, showing the proportion of correct response is typically -- are greater than or equal to 90 percent and has been consistent across all survey waves.

This finding was similarly seen when we assessed MS infusion site knowledge.

We also asked questions around knowledge, attitude and behavior in our TYSABRI surveys, in particular attitude and behavior.

And some of the results are shown here. Most prescribers agree that the TOUCH

345

program is the appropriate way to monitor the safe use of TYSABRI. While 93 percent agree that they have adequate resources to complete the TOUCH program requirements, 20 percent indicated that they had to hire new personnel or redirect existing personnel to fulfil TOUCH requirements.

Twenty-seven percent were uncertain if TOUCH program may prevent access to TYSABRI for some patients because of practice administrative burden.

But we believe that the current survey methods adequately assess if a program meets the REMS goals of informing and educating stakeholders about the risk of a drug.

As we have heard today there are limitations to those survey methods. Importantly there are limitations including lack of assessment of in-depth stakeholder knowledge and views regarding understanding of benefits and benefit risk balance,

effectiveness of individual REMS tools, healthcare system burden of REMS, and barriers to access.

The stakeholder surveys include only those currently participating in REMS, and therefore are not adequate to fully understand barriers to access or burden.

We are investigating alternative approaches to REM surveys, including assessing stakeholders' understanding of benefits, assessing stakeholders' perception of benefit-risk balance in REMS communication and REMS tools, and surveying eligible stakeholders not currently participating in REMS to more fully understand barriers to access. Thank you very much.

MODERATOR TOIGO: Thank you, Dr. Cristiano. Questions for Dr. Cristiano? Okay, then thank you to all of our speakers in the open public session. We appreciate you taking the time to prepare your comments and then to come and deliver them. They do

enrich our process.

So thank you. So now we are going to get to topic 5. Julia will put the question up, and I will read it, just for the record.

We are going to -- these are alternatives to surveys. Given the issues with surveys that we have observed, what are the alternatives to knowledge surveys to assess the effectiveness of the educational elements of the REMS?

If any, what are the advantages and disadvantages of the alternatives?

Who wants to start our discussion with this question?

Okay. Please.

PANEL MEMBER GILSENAN: So, I'd like to get a little bit of clarification on what exactly you're looking for. When I first read the question, my immediate thought was to do a database study to see if the prescribing behavior or the utilization in

1 something like that, but I'm not sure if
2 that's what you are asking about -- it's the
3 effectiveness of the educational elements of
4 the REMS. So what exactly are you referring
5 to there?
6 MODERATOR TOIGO: Before I comment,
7 I'm going to pass it to the side here and let
8 my colleagues address that.
9 DEPUTY DIRECTOR SHIBUYA: Rob
10 Shibuya. I think this is going to be a
11 difficult question to answer for the panel,
12 so I'm going to ask you initially to be very
13 limited and think about the goal that Dr.
14 Willy said: it's mainly usually
15 informational, and some alternative to a
16 survey that would directly inform that goal.
17 And then if that is not
18 particularly productive, I think we could
19 broaden the conversation. But I was hoping
20 to initially get a very -- really something
21 that directly informs whether or not patients
22 and prescribers are aware of the important

risks.

MODERATOR TOIGO: Dr. Lefkowitz.

PANEL MEMBER LEFKOWITZ: One alternative to a survey that still has a data collection element would be to design some sort of educational intervention for patients, something web-based, something where they basically would review the items that are in the Medication Guide or the other things and have their knowledge immediately assessed.

That has advantages in that the patient doesn't have to do something else. They are not being contacted out of context. They are receiving the information at the same time.

You could ultimately decide if the educational intervention was sufficient; you don't even have to test them. Yes, like programmed learning, like what's done for a lot of things that people have to do on their jobs, to learn about X, Y or Z.

I mean, there are models of adult learning and how to assess adult learning out there, that don't involve surveys.

MODERATOR TOIGO: Okay. Thank you. Dr. Moratto, were you going to comment? And then Dr. Shiffman.

PANEL MEMBER MORATTO: Yes, so if we kind of open up the idea, brainstorming, I kind of look at this as -- you are collecting a mosaic of evidence and you are looking at the totality of evidence, rather than saying one definitive trial or one definitive survey.

So in that context, I think qualitative methodology with the key stakeholders, as we have just heard from many of the open forum speakers, with target groups of who are the initial adopters.

So you could imagine, when a drug is first approved, you have some idea which practices or clinics are starting to use it first, and do guided, focused interviewing on

there to understand what their knowledge is, and maybe a series of these over time would be a way of, as you are trying to understand, as it's going from the high user, early adopter and spreading out more broadly into the population. So that might be one methodology.

You know, with qualitative, it's not large sample necessarily, so the generalizability goes down, but it would help you understand the whys that you are not able to look at in terms of quantitative testing.

Another approach that I think could be considered would be what we might call directly observed behaviors. There's methodologies that have been used that look at the provider-patient interaction.

One that is used is the Roter interactional analysis system. Others have done other direct observation of primary care studies -- Dr. Stange, for example, in Ohio.

So the idea is you are now -- you

are measuring the interaction that is going on between the provider and the patient as a way of -- again, it's the application of that knowledge, and to what degree is that information being transferred from provider to patient in that setting, would be another way to complement, perhaps, surveys.

And then, I don't know if you mean surveys in the context of how we talk surveys, or you just mean all closed-ended questions? I mean, because even what you are describing will have a closed-ended question, or do you mean just surveys as it was kind of constructed this morning? Because there's other ways that you can be getting that kind of information: as an intervention. There can be work that is done directly at the pharmacy site in which patients, when they are -- if you have a drug that is being used often enough, you'll have some sampling, you can do sort of point of surveying or interviewing, or cards handed out.

So it falls within that kind of data collection, but the way you are getting a sample is a little different. So I wasn't sure how you were construing survey.

DEPUTY DIRECTOR SHIBUYA: Rob Shibuya. I think we were construing survey in the way that you have heard, that most of the ones that had been submitted to us, what those -- the methodologies that they have been employing.

PANEL MEMBER MORATTO: So, if -- I don't know if it fits within it then, but I think one of the sponsors brought up the idea that you have patient user groups. I mean, that would fit with the idea of the qualitative kind of study in which you have maybe an online patient user group or some other forum, and you are now either questioning surveying them or you are using that as your sample for qualitative discussions and so forth. So I would add that too.

354

So again, all of these, you lose the ability to say it's a nationally representative sample, it's been -- you know, the strengths that you have with the survey, but you're -- it's a way to get sort of the mosaic of many pieces of this so you have an understanding over time, you know, what people are thinking.

MODERATOR TOIGO: Thank you Dr. Moratto. Dr. Shiffman then Dr. Wolf and Dr. Day and then Dr. Davis, so that end of the table.

PANEL MEMBER SHIFFMAN: I think we touched on this in the earlier panel, but I think it bears repeating that even prior to going out to seeing whether the level of knowledge in the population of actual patients using the drug, it seems a very important, at least necessary, if not sufficient, first step is to see whether the materials that are being used are adequately comprehended.

355

And, again, there are established methods for doing this covered by FDA guidance. One of the big advantages of that approach is that you can test before launch. The drug doesn't even have to be known to anyone. You don't necessarily test in patient populations, but conversely you can be quite sensitive to different sub-populations. You could do a test in people with limited literacy. You could do a test in the elderly. You could do a test in people who are already using other drugs and so on.

I think those methods, as covered by current FDA guidance, are all verbal. They are based typically on scenarios, and the responses are open-ended verbal responses, they're face to face in a sense qualitative studies, but there's no reason they need to be limited to verbal responses.

If, for example, an issue is proper disposal of a medication, one could ask a respondent to demonstrate what they would do

with the medication, or if an issue is appropriate administration.

So it's more of an assessment in principle, but it has a lot of flexibility. And, again, it's not necessarily a substitute for assessing actual knowledge, but if you know in advance that the materials don't communicate adequately, distributing them widely and then assessing the effect seems rather pointless.

So this seems like -- it's almost like clinical trial done prior to approval and then you have post-marketing surveillance. This is the pre-approval clinical trial.

So in a sense it's a substitute for some of what you want from the survey, and the focus is really on the effectiveness of the material.

And if you think about it, if the material is effective and then you are able to show in the world that the materials are

actually distributed as you intend, and consumed, if you will, by the patients, you have pretty good assurance that you are going to disseminate knowledge because you know the materials work when they are read.

So I think those need to be a very important part of what I agree needs to be a kind of totality of evidence approach.

MODERATOR TOIGO: Thank you, Dr. Shiffman. Dr. Izem is going to comment and then we'll go down to the end of the table.

DR. IZEM: Yes, I wanted to make a quick comment. Were the studies that you were mentioning sort of consumer behavior studies for over the counter drugs?

PANEL MEMBER SHIFFMAN: Yes, that's very much what I have in mind.

DR. IZEM: Okay. I just wanted to make a -- I guess an important distinction. If you are going to purchase an over the counter drug, you are not going to talk to a learned intermediary, so you are supposed to

understand the label without a learned intermediary.

So then all this testing that you are saying makes sense, however, if you are -- if you are an approved drug that is a prescription drug, you have a label and you also have to talk to your doctor. So it's not -- the Medication Guide or the label is not the only information that is available to the patient, and sometimes the terminology that is used in the Medication Guide, you are supposed to talk to your doctor, who may boil down the information in a more understandable language, I guess. That's what -- I wanted just to make that clarification.

PANEL MEMBER SHIFFMAN: I very much appreciate the distinction, but let me argue that too much is made of that. If we thought that the prescriber was doing that effectively, the Med Guide and other sources of information wouldn't be necessary.

The point is that we are trying to

supplement what we hope the provider is doing, and that once the patient gets home, it may not be an OTC product but it's left to the patient to use it appropriately, to recognize treatment, emerging symptoms or conditions and so on.

Now, importantly, this is true also for an OTC, but perhaps doubly so for a REMS-covered prescription drug, one of the options for the patient who has a troubling symptom is to call their provider or prescriber or pharmacist, and that would be I think considered a very positive step.

So a respondent who said, you know, if that happened I would, or the person in the scenario should call their doctor, we would want to consider that to be an appropriate response.

But in the end, the purpose of Med Guides is to supplement what we hope prescribers are doing.

MODERATOR TOIGO: Thank you, Dr.

Shiffman. Okay, we'll start with Dr. Wolf and then just go right down the first four people.

PANEL MEMBER WOLF: Great, so I guess I'm taking maybe a very -- when I first read this question, and it is kind of difficult, and instead of thinking of -- I thought maybe instead of just thinking about the limitations that were talked about in the first panel about surveys, and then what the outcomes would be in reviewing those challenges, that maybe the other hidden question, and I really like, Elaine, how you said this is a mosaic, that you are trying to get as many pieces of information that you can to really try and get an understanding and rich assessment of any REMS, that it's also about being efficient and finding ways, whether it's data that is going to be increasingly available that you can use as proxies for the quality and effectiveness of REMS guide, whether it may not be as direct

as what you would get in a survey.

So -- and I also agree with Saul that you know you have comments where you start looking at the usability of materials, you can do things that just are -- I think I'm assuming you are already doing anyway or that you should be doing anyway with regards to critiquing some of the existing Medication Guides beyond getting at comprehension and surveys.

But also even looking at usability of medication devices, because I think a lot of these assessments are dependent, and where you get data and what kind of data you look for are going to be dependent on what's included in a particular REMS.

So is the important message that is being conveyed in a Medication Guide is how important it is to adhere to a medication, like for some HIV drugs, or is it about safety, like for all of the REMS?

But for fidelity of outreach to

providers, I was just, you know, whether -- I don't know how you disseminate the materials, but are there ways of getting automatic or electronic confirmation that they at least received the Medication Guide, because in some work in the past you find that just even getting the -- not the Medication Guide but the REMS communications -- is difficult.

And so if there's any resources where the ways you disseminate the REMS to the provider's standpoint is the fidelity assessment, that you can find ways just to get, you know, large amounts of data, then at least you know that it's being received.

And then a lot of what we are doing with electronic health records, both on the pharmacy records side and on EHRs, as with meaningful use, you see them continually rolling out, how can you make sure that there are new fields?

And I guess I don't know, this might be a question to confirm, I'm pretty

363

sure from our system we don't, on the prescriber's side of it have a current field that is confirming that a Med Guide was given to a patient.

You know, so again, this is supplemental information that you could get lots and lots of rich data. I mean, we are not talking about qualitative which would be great too, but on the other scheme of it, tons of data from health systems that would be confirming certain prescriptions that they are at least receiving the medication.

And then I think it gets harder of course when you start looking at outcomes dependent on medication. Do you want to ensure that, if adherence is part of the REMS for medications, the message -- that data is probably more readily available than getting information about safety information and I think that's probably still be thinking about.

And one last comment, to be

efficient here, is secondary data also -- the Consumer Assessment of Healthcare Providers survey, the CAHPS questions, actually do include a lot of assessments from consumers about what is being said to providers -- I mean, what providers are saying to patients.

And I know how many health insurance plans and health systems are using CAHPS. I don't know if there's a way that that could also be built back in, or maybe even -- and again, I feel like it's kind of a cheating answer, because it is technically a survey, but it's another existing secondary data resource where maybe you want information, because it does say things, did your provider talk to you about risks of medications and so forth, with some of the more recent build-outs of CAHPS, so --

MODERATOR TOIGO: Thank you, Dr. Wolf. I am going to ask my FDA colleagues if you have any questions that you wanted to pose to Dr. Wolf based on his comments.

365

DEPUTY DIRECTOR SHIBUYA: I'm trying to take good notes here. So can you tell me what CAHPS stands for? And I think Dr. Moratto you said something like Roter something at some point and I wanted to capture that.

PANEL MEMBER MORATTO: Yes, Deborah Roter, R-O-T-E-R. And then she has a -- she's out of Hopkins, has a system, Roter interactional analysis system.

And since I have the mic can I just add one thing real quick, because you made me think of something else.

So as you think about available data sources that's really smart, and there is so much that is being invested right now in terms of comparative effectiveness registries, the patient-centered outcomes research registries, and also within the NIH has these community translational organizations that have community engagement corps that are tapped into the practice-based

research networks.

So these folks all -- there's many of these networks nationally, they all focus on different probably disease states, some more than others, but it's one way to go in-depth within a health system. Instead of breadth you are going really in depth and I would imagine many of them are interested in the same kinds of dissemination implementation science kinds of questions.

PANEL MEMBER WOLF: So, just to throw into the pot, the CAHPS stands for, I believe, just if anybody wants to jump in, Consumer Assessment of Healthcare Providers Survey. And again that's been used as a way to kind of at least rate quality for many healthcare plans and so forth, and I think it's got a lot of uses.

And one last thing, and this may be kind of really being out on a limb here for secondary data, but I remember at one of the last risk communication advisory committee

panels, Nancy, that we had, the MedWatch program, which is now, isn't it going online?

And I don't know how this could be used, it's just a thought that now if consumers are going to be increasingly available to report symptoms or side effects associated with any taking of any medications, I think, and I feel a little bit like I'm not -- a little fuzzy as I'm talking about it. But that could be one more way I'm thinking that you might be able to find links to patients that, you know, at least follow-up opportunities if it's linked to something that's in a Med Guide.

MODERATOR TOIGO: Thank you, Dr. Wolf. Dr. Day. Oh, no, Dr. Davis.

PANEL MEMBER DAVIS: So, one point is we have the possible and the practical, and so these data sets are practical because they are there.

In the possible, just to brainstorm, if I was the Queen and this was

my problem I would start qualitatively. Clinically, in teaching medical students in residence, I would ask them to do a teach-back with a patient to confirm clinically that they had understood.

To Saul's point, that we wouldn't need this if they were conveying the information, doing the teach-back and the patient got it when they walked out of the pharmacy or walked out of the physician's office.

This was just a brainstorm. Several years ago we did a study on childhood immunization. Doctors weren't giving risk information. We made a poster to put in exam rooms: seven questions parents need to know about baby shots, and that was the teachable moment. And in studies we did, it increased physician, nurse and parent communication about the essential risk-benefit communication of childhood immunization, and WHO has just said they want to have that as a

best practice for Europe next year.

Oh, and one other thought. You know, all this web-based stuff is so cool, but then, in Mike's words, you have got the skill set and the mind set. So, many people would not have the skill set to do that. I have the skill set, I don't know if I'd have the mind set to do it.

MODERATOR TOIGO: Thank you, Dr. Davis.

PANEL MEMBER DAY: I would like to remind everyone that there are some cognitive paradigms for testing knowledge and a variety of other kinds of things that have been around for decades, and these methods have been out and tested and developed and they can be used to find out not only what people know but how they know it, and then we can use them to find out why they know it in a certain way.

So some of this has come up a little bit today but I'd like to reword it

and introduce some concepts that might be useful to everyone.

One is levels of processing. If there is a key risk message that we want to across so people understand and can use and apply, that testing it in multiple ways from free report -- now this isn't open-ended anything anything, that as Dr. Ostrove has pointed out, then requires a lot of coding and burden on data analysis and so on, but it can be constricted to a smaller content area.

Free report, then give a little cue to get more, that's cued report, then go into a recognition paradigm. We had two examples of recognition paradigms that have come up repeatedly here today, true/false and multiple choice, but there's lots of others as well.

And with these -- so you would have multiple testing within -- for the same content, and using these approaches you will find that chance varies. Chance came up just

a little bit today. So when you are trying to set a knowledge rate or whatever, no one has brought up the fact that it depends on the nature of the question.

So if you have a question with two possible answers, yes or no or true or false, chance is 50 percent. So if you get a percent correct at say 65 percent for that, as opposed to one of the ones where chance is 25 percent, because you have four alternatives, it has a different meaning.

So looking at these cognitive paradigms helps you interpret what you are already getting in the first place. Secondly, it helped -- we didn't hear anything today about an error analysis, only looking at what do people get correct. Now, they can get it correct partly because of chance, but when they make an error what kinds of errors occur?

Going back to the lowly true/false for just a moment -- it's got some value but

it is more lowly relative to other things. When somebody gets it wrong, what is the it they got wrong? It could have been when the item was true and they said false, or it could have been when the item was false and they said true, all right?

And maybe people are doing this kind of analysis, but we didn't hear it today. And by doing this you can also assess -- I mean you know then what the difficulty is in the understanding and can take steps to make it better, but you can also see whether there are some people who are naysayers and some people who are yeasayers.

So as you take this idea of error analysis up from the true/false and go to the recognition, say, in a multiple choice, four choice alternative, there are then more things that you can look at. You can find out about discrimination between details in the information. You can find out about bias and so on.

So there are things called hits and false alarms and there's this whole science of signal detection. So you can find out not just that we got, say, 75 percent correct on something, but what was going on, what was the nature of the errors.

Another concept and then I will quit is the idea of a problem space. So there's been a lot of positive things said about testing scenarios or vignettes.

So if, say, you want to understand whether people know, whether they're the providers or the patients, whether the drug is appropriate for a given person, you can say, okay, this person is 85 years old, is it all right to take it, or whatever you are going to, but there might be three criteria or more. That's one of the problems that the statins have had in trying to get a switch from prescription to over the counter, because of the self-selection problem.

Well, here now we have prescription

drugs where the provider is supposed to know all of this, and the patient is supposed to understand to a certain extent.

But what if the patient goes out and has been prescribed something and then gets a new health condition? All right? Does the patient even think to talk to the doctor about it, and it may be the drug might not be right for them anymore.

So in order to look at, say, an indication where there is like three criteria, that means that there is actually eight scenarios that could occur within that space, because you have got three different criteria, you know, A, B and C, you know, one might be age over/under something, et cetera, so for each one you could fulfil or not fulfil that criterion for a given patient. Taken all together you could have -- fulfil one of them, any one of three, two of them, three of them and so on.

So before -- you may not want to

test all that, but before you go and pick out one of them, you have to know what the whole problem space is. Are you asking the easiest one? Are you asking the hardest one? Are you going to be able to really tell whether there is understanding and ability to use?

So there are a lot of other issues here about who to test. Should they be non-patients and early versus late in development, early testing, these cognitive principles can go all the way through from the initial pre-testing all the way to the final testing.

So in the actual REMS the cognitive paradigms can be used and can be used successfully. We do it all the time in my lab, from people who are healthy, have no health conditions that we know of, to people who are likely to be prescribed certain drugs versus those patients who are actually on them.

So these principles, although I've

emphasized some of the application to the knowledge domain, these principles are important for whatever testing can be done, whatever paradigms, the new and cool ones as well.

MODERATOR TOIGO: Dr. Wolf, did you have your hand up because you wanted to comment? Okay. Before we go to Dr. Eggers. Okay. But before we go to Dr. Eggers, any questions from the panel for Dr. Day or clarifying? Okay. Thank you, Dr. Day. Dr. Eggers.

PANEL MEMBER EGGERS: Thank you and I'll preface my comments by saying I'm sure glad I'm not going to be asked to follow up on any constraints with anything that I am going to say today or the constraints of actually doing this.

But I would like to follow up on Elaine's comment about the whys of the behavior and I would call this sort of the context of the decision.

This is sort of another preface to what I'm going to suggest is maybe something to think about methods. I think there is value in understanding that context of the decision as much as understanding the knowledge, which is one of the whys of behavior.

And I'll just point out I think there are three that I can come up with off the top of my head but there are likely more. One is prior beliefs, things -- that can include things like credibility, et cetera. Values and other decision-making criteria, so what do people take into account besides their knowledge, objectives, tolerances, risk tolerances, tradeoffs, et cetera. And then finally what I'll call the system influences, which are all the other things that can influence our decision: time, access, ability to take the bus to get to the doctor, et cetera.

And if those things were talked

about as endpoints this morning then I apologize because I wasn't on there, these could be sort of sub-endpoints one could think of.

And so with those two prefaces in mind, I think that there are some criteria that could be used when thinking about the methods, and perhaps bringing it out of the context of just one specific REMS, one specific drug, and broaden it. Is there an opportunity to look at broadening across, say, a drug class or across things that would have the same sort of characteristics of risk?

So that could -- and by doing so you could develop say modules of decision-making and maybe help take lessons learned that even if you can't apply to a particular assessment, you maybe could understand a bit better what's happening with one particular drug or give general guidance.

I think early, if there is a way to

make those methods so that they are not just when the REMS is in place, and at the assessment phase, but methods that can be used early, maybe the more qualitative ones first, in the design phase.

Mixed methods, I think -- so from my background I have used a method called behavioral decision research which combines normative decision analysis with the qualitative that everyone is talking about, exploratory sort of decision-making, have someone walk you through the decision they are making, followed by the assessments and followed by more structured surveys.

And so I think there is a real value coupling the early with the mixed methods, the decision analysis type tools, and I think FDA might be exploring those already.

And with that, then, I think there's iterative and building, and so building upon that, knowledge assessments

don't have to as much start from scratch or be so constrained. They can draw on what was learned before. I think Saul was making some points about that as well.

So I'll end with that preface again that I didn't have to think about any constraints to doing this. So thank you very much.

MODERATOR TOIGO: Thank you, Dr. Eggers. Do any of our industry representatives want to comment at all on the last four comments that you heard? In terms of practicality or just -- if no, then that's fine but I just wanted to open it up if anyone wanted to comment. Okay, well, you can -- we'll start -- let Dr. Morris go and then we'll -- Dr. Peterson.

PANEL MEMBER MORRIS: I guess a lot of this recent discussion is really, I think, very important and very interesting from the standpoint of developing communications and understanding what people know, and getting a

very broad assessment.

And what I'm trying to do is trying to figure out where this kind of fits in in terms of a REMS assessment and I'm having hard time because I'm -- part of the problem is that -- and this is probably -- this question may be a really good reason why we should pre-test questions, because I'm getting very hung up with this -- kind of parse out what you really want to know here, and I'm hung up on this phrase knowledge surveys, and the -- I think the question is saying, other than asking people what they know, how can we measure what they know? Is that what you're asking?

MR. GORDON: Yes, I think that is a fair statement. If we are not going to use surveys to assess knowledge, how can we measure what people know? Are there alternatives we could --

PANEL MEMBER MORRIS: Yes, so other than asking them -- and I don't know of a

good way. I mean, personally, I mean you can observe behavior, you can kind of observe the input and the output but the little black box of what's in the brain is hard to assess.

I could see you can do functional MRIs and figure out what part of their brain gets lit up, you know, when they read, you know, a Med Guide, and there's things like that you can do but in terms of -- I don't know if we are at level yet if we can know, you know, to actually identify physically a mental model where, you know, what neurons kind of light up when we -- you know, when we -- but other thank asking them, it's hard. I think there are maybe innovative ways of asking. I think at lot of what Ruth was saying was, you know, kind of you can do it in innovative ways and I really liked the kind of the mental models approach and, you know, rather -- you know, you can kind of try to measure their concept of what they know rather than trying to measure do they know X

or do they know Y, you know the way we are doing it now.

I always thought copy testing was a great kind of model for that, you know, where you actually ask them to tell you back what it means and ask them very open-ended questions in a very careful way.

But it doesn't tell you if they meet a threshold. I mean, if your -- if your job is how do we know if a REMS is working or not, and you're going to have to establish a threshold, then, you know, you are kind of locked into having some kind of question that, you know, you can answer that is either correct or incorrect, and develop a threshold.

But, you know, again there's innovative ways of doing it but most of those innovative ways apply to product development or information development as opposed to an evaluation of whether something meets a certain threshold or not.

MODERATOR TOIGO: Dr. Smith, you were going to comment?

PANEL MEMBER SMITH: Oh, my comment just was in general to the whole idea of using a multi-method approach to collecting information about the impact of REMS. And we touched in that in our industry panel but we really endorse that. I mean, that's a really great way to think about it.

MODERATOR TOIGO: Thank you. Dr. Holmes, were you going to comment? And then Dr. Lee.

PANEL MEMBER HOLMES: Well, I want to just add that while we endorse that, we think there's lots of caveats to endorsing that, key one being that many of what these multi-methods speak to is the -- is the qualitative work that underpins what we are asking people, if we ultimately choose to go and ask them what their knowledge is, and not all of these things apply to every REMS. So, yes, we endorse multi-methods, but we --

there's a lot of caveats to that endorsement, for some of us at least.

MODERATOR TOIGO: Okay. Thank you, Dr. Holmes. Dr. Lee and then Dr. Wolf.

PANEL MEMBER MATTHEW LEE: Sure, I just want to re-investigate the idea of using a different data set to evaluate the effectiveness of the REMS and the effectiveness of the educational elements.

One of the things that we strive to do is marrying our post-marketing safety data with our REMS evaluations, and, you know, let's say, for instance, a particular drug causes some sort of liver malfunction, and does your post-marketing, you know, safety evaluation database show that there are a number of reports coming in of elevated liver enzymes, or does it show a large number or a large influx of full-on liver failures?

You know, are we catching the safety signals early or are we catching them late? Is there a way to marry the post-

marketing safety surveillance data that pharmaceutical companies are required to collect and report to the FDA with the risks that we are trying to mitigate?

MODERATOR TOIGO: Dr. Willy or Dr. Shibuya, do you want to comment?

DEPUTY DIRECTOR SHIBUYA: Well, I think in a lot of our REMS we do also take the same sort of multi-modal approach. We try to look at the effect of the REMS and the overall safety of the drug in many different ways. So I think we're sort of on the same page.

MODERATOR TOIGO: Dr. Wolf.

PANEL MEMBER WOLF: I'm sorry, I'm actually, due to a poor-planned flight, I want to get my last couple of comments to throw in. You know, to that one point, I would just kind of like to supplement on the strategy where you are getting proxy data or you are using existing data to kind of complement, and again, I

stress the word complement. There's so much that if we were talking with -- again, I'll stress right now, you know, EHR vendors, pharmacy record vendors, and often they're one and the same because they are creating multiple products, that you can not only turn surveillance -- you can turn surveillance of REMS, at least in some ways, ultimately into interventions, where almost REMS are being built in, the communications themselves of REMS, are being built into the systems because you are putting prompts for a provider to either convey or deliver a piece of information and get that piece of information confirmed.

But also you can make sure that at regular intervals that certain elements are being evaluated that you need to, as well as the fact that you are now going to be able to merge perhaps behavioral data, knowledge data, or demonstrable knowledge data, as well with clinical data.

So again, whether you be looking at a lab value or -- and hopefully you are not waiting three months to find out that somebody's clinical outcomes were kind of below goal and that was part of the REMS, or that they experienced some sort of adverse event.

But you can get real-time information quite -- I would imagine that this stuff is no longer just outside of reach. It's in reach. It's just getting the conversation started.

And one completely tangent that I wanted to throw in here, the other thing that -- because I'm really focused in on practical strategies, things that you could possibly be doing tomorrow while you continue to have these discussions about the optimal knowledge survey, the best way to do it, the timing of it, really, really big picture and difficult questions, that you could be also -- you know, if you target high-risk groups, groups

that are likely to be taking these medications, or just be taking medications in general, with ongoing initiatives like medication therapy management, so MTM services, how could you think of REMS information being collected? Home health visits, people leaving the hospital or, you know, could you have assessments that would get at information, whether it be from storage content of a medication, to information and demonstrated understanding of certain side effects, how could you capitalize on activities that are going through already, with checklists or inventory lists in addition to the stuff that you might be collecting with, again that would be new fields that you would be populating in a medical record.

MODERATOR TOIGO: Thank you, Dr. Wolf. Dr. Holmes.

PANEL MEMBER HOLMES: Can I just take my industry cap off for a second and put

my doctor hat on, and warn against -- and I love electronic healthcare records and they are wonderful, but more and more and more checkoffs make more and more of us doctors more and more checkoff artists.

And so you know, it makes it hard to actually take care of a patient and talk to them and actually have the interaction with them. So I guess I love-love those ideas, but they also have to be balanced with what happens in the room, and the relationship with the patient and making sure we are not creating too many checkoffs that satisfy our REMS but actually undermine the patient-provider relationship.

MODERATOR TOIGO: Thank you Dr. Holmes. Dr. Hornbuckle.

PANEL MEMBER HORNBUCKLE: This is a very, very rich discussion on alternatives to surveys. I think this is very good that we are having this discussion.

One important thing is to make sure

that alternatives also are aligned back to the REMS goals and the objectives. I think that's a very important point.

And then two, hearing all these alternatives, one needs to also think about the practicality as well as feasibility, and I mention that because clearly with the REMS requirement, there's a timeline. There's a time frame that we have to work in.

Hearing many of these alternatives definitely would be a challenge in meeting these, the REMS timelines, the 18 months. Some compounds have -- some medications have six months annual REMS assessment, so we also need to think in terms of the practicality as well as the feasibility in order to consider these additional approaches as well.

MODERATOR TOIGO: Thank you. Dr. Moratto and Dr. Lefkowitz.

PANEL MEMBER MORATTO: I agree. And I think we could think about -- if we go back 50 years when pre-IND and phase 1, 2 and

3 were being conceived after the Kefauver Harris Amendment, we are kind of at that same stage of thinking, but thinking risk management, right, what's the phases of risk management development.

And I would argue that we should be thinking, what can we be building in as we are doing clinical development programs so you may not know if you have a REMS, but you know what safety issues are coming down the road on the drug.

And we should know what's current habits and practices of practices and how they handle those kinds of things. I'm sure the sales organizations understand the practicality of how the doctors are practicing on patients and those schedules so that there's formative work that I think can be done during the development of the drugs in parallel to the clinical trials, and this can be qualitative. It doesn't have to be high-expense kind of thing, but gives you the

foundation so that when you have this narrow window and you are suddenly asked to devise, here's the education, we are not all starting from scratch.

But I would agree that I think part of the challenge we have here is -- and maybe it's the regulatory constraint -- is to do something well, you can't be designing something in three months. You don't develop that marketing materials to launch the drug in three months from the time you have the labeling to you launch. It was going on before that. There was some anticipation and it may be refined and finalized and shaped at the very end, but it's not starting from scratch.

And so as I sit on the drug safety risk management committee, and we get to review drugs when they are at the advisory committee stage, wanting to get approved, it's amazing how little information is being shared as to any formative work on current

understanding of their knowledge of risks that are pertinent to the drug or to behaviors.

And I think we should be -- FDA can be encouraging more of that input is done earlier in the process and that should be part of the piece that is forming and shaping the REMS, not having you need a REMS, now let's go develop education materials.

And I know that puts more burden back on drug developers and that, but I think if we think of this as a life cycle approach, which is where the IOM report's going and previous reports, you know, what can we be doing earlier on, how do we take advantage of the market research data that is being collected to also be piggy-backing on and getting more you know, the drug safety awareness and behaviors too.

MODERATOR TOIGO: Thank you, Dr. Moratto. Dr. Lefkowitz.

PANEL MEMBER LEFKOWITZ: I want to

1 challenge one of the assumptions that's been
2 put on the table, which is the best way to
3 mitigate risk is to ensure that the patient
4 has these specific bullet points of
5 knowledge.
6 I mean one thing that Dr. Wolf
7 mentioned earlier was the CAHPS surveys which
8 AHRQ developed, and what the CAHPS surveys do
9 is they assess patient satisfaction without
10 ever talking about are you satisfied with
11 your doctor.
12 What they did is they took elements
13 that, from the patient's perspective would be
14 considered a satisfactory experience, and
15 then they asked about that, rather than
16 saying, are you satisfied.
17 You could imagine the same sort of
18 thing being developed here where you decide,
19 you know, what are the aspects of patient-
20 provider communication or the whole package
21 of information someone gets when they are
22 deciding to take a drug with serious risks,

and say what are the questions that I need to ask to ensure that those communications took place, rather than asking about the specific knowledge that came out of it.

One advantage to that is that you don't have to reinvent the wheel with every REM. That you could imagine a database of questions being developed that would assess that communication pattern. You could think about adding those to existing REMS which might be more, you know, knowledge-specific, to see in the long run what predicts what.

I mean, I think you know that we are sort of hung up on how do I get people to know this one little factoid. Is that really what makes somebody safer over the long haul? Or is it assessing did this communication happen to ensure that the person got some information at the beginning and knew what to do in the long haul?

MODERATOR TOIGO: Thank you. I thought there was a hand. No, no hands.

Okay so let me put it back to the panel here and see if there's additional discussion or questions that you want to get some input on.

Okay, so then we'll move to the last topic, which is assessing behavior, burden and access and this is another question. What are the considerations in designing questions to assess the impact of REMS on patient and/or provider behavior and access to drug, as well as the potential burden of the REMS on these groups.

What are alternative methods to assess behavior, burden and access for a REMS? If any, what are the advantages and disadvantages of the alternatives?

Dr. Shiffman and then Dr. Moratto.

PANEL MEMBER SHIFFMAN: I want to really emphasize a point that was made by one of the public comment speakers, which is that you can't assess the burden by assessing the patients who actually got the drug and the providers who enrolled, if it's a provider

registry program, because you are assessing those who succeeded in getting access and it's very important to understand the patients who gave up because the burdens were too great, and the providers who, whether because of the burdens of that REM or here's where the cumulative burden of multiple REMS comes in, where the providers say I'm just not going to do this anymore.

Now that increases the difficulty already discussed this morning. We said it's hard to get even the patients who are on the drug, but I think to assess burden we need to look at the patients who perhaps were prescribed the drug and never filled the prescription, again, administrative medical records may enable you to identify those and perhaps more difficult, to assess providers who might otherwise have participated and who passively opt out because the burden is too great.

And again here, I think on the one

hand, the burden of REMS by statute and practice is on the individual sponsor for an individual drug, but I think the agency has to look at the cumulative burden because what we hear from prescribers is that it's not any one REMS, it's the cumulative burden that is making them opt out.

And that does suggest that there needs to be more coordination, you know, category-wide REMS common practices that would minimize the burden on prescribers.

MODERATOR TOIGO: Thank you, Dr. Shiffman. Dr. Moratto.

PANEL MEMBER MORATTO: Yes, I would think the idea I would like to add would be maybe some approaches that come from systems engineering or industrial engineering, in which you might, it's -- the scenario of the patient, the practice, all the people in that chain of events and what are the steps that they have to do in order to go from point A of getting the drug to point B, and doing a

good understanding of what's happening in there.

So for instance, I'll -- you know, so these are like time and motion studies. I know they have been done in pharmacy, they have been done in industrial sites, et cetera.

But this -- these methods are also being applied towards healthcare, and how do we think more efficiency in the delivery of healthcare using these modeling methods.

So for example, to illustrate, if we were to think of isotretinoin-containing products, and I have a patient who is an 18-aged girl, who is referred from their pediatrician. That's where I have the patient relationship.

Now, they have to find a doctor that's part of the REMS program, so you can say okay, where's the proximity for that. They are going to have to establish the relationship with that doctor. They are

going to have to be able to get prescribed a birth control pill. That might require a different doctor, may not be the pediatrician, may not be the dermatologist. You know, they are not going to have to get blood draws, what practice can they go to get that, and you start to create the scenario of what are these steps, and then there's modeling that you can do to say, well how common is this particular scenario versus that, and you can run through the system.

And you might do some thought experiments to say, well, what was the old way, if it was a drug already on the market, or what's an alternative way of doing this versus another, and you'll get some sense there for a number of steps, time involved in a practice, you know, if they are having to hire someone or dedicate a certain amount of time to a nurse or practitioner in the office to follow through on patients, you can put cost estimates on those, right?

But it's a way of thinking of it as a system of all of these affected stakeholders and how are they affected by the REMS. And ideally you might be able to use this in a cost effectiveness-like modeling that says we have called it a cost of dollars or a cost of time, of steps, what have you. What is the cost effectiveness of that per bad event avoided that you are trying to deal with, you know, bad liver failure or heart attack or what have you.

And you might therefore start to build some science around, just like with qualities maybe down the road, what might be a threshold where we say yes, there's more burden, but we also are trying to protect people and there's probably a certain tradeoff of burden that we are you know, we are willing to tolerate versus not, depending on what the adverse affect is, right?

And I think we run the risk, as we sit and cross, you know, as you are saying,

multiple REMS are going out, and they have similar risks that they are trying to manage.

Let's say for instance around teratogenic risk and you now have different drugs have different teratogenic risk management programs.

The next logical question is going to be well why does this drug have this level and this one doesn't have this level, and you are going to need some way to compare you know, what is the benefit of this added incremental level of risk management in terms of the elements of safe use.

And this might give you some ways to start to quantify or to start to think about how you might quantify that as an approach. So, just share that.

PANEL MEMBER SHIFFMAN: Just a quick comment on that. I agree very much. But to bring it into, under a rubric that FDA is already using, for example in DMEPA, would be FMEA, failure modes and effects analysis,

404

that similar sort of systems analysis to kind of -- you could brainstorm it in advance and you could collect data about how it's going, and consider that there are two modes of failure. One is that a patient gets the drug who shouldn't get it and is put at undue risk, or isn't informed appropriately, but the other form of failure is that a patient who could benefit from the drug doesn't get access to it or pays too high a cost, including their time and hassle.

And to then -- it doesn't make it a simple problem, but then at least you are weighing the cost to many against the cost to few who might be harmed.

MODERATOR TOIGO: Thank you. And there are discussions certainly internally around a lot of these issues. We are trying to, you know, this is just the assessment piece, but many of the things you are raising, we have people internally looking at those and trying to grapple with how do we do

this. So thank you. Dr. Day and then Dr. Gilsenan, Davis, okay.

PANEL MEMBER DAY: When we talk about behavior change, it implies two measurements at least, before and after a REMS would be in place, and there must be data on hand that can be looked at.

As a former member of the DSaRM Committee we had several meetings on isotretinoin before there were various risk maps and so on and so forth.

So you could see that, say, prescribing would drop after one of these sets of measures would go in but then it would come back up and then more would be instituted and so on.

So there's a wealth of data out there that could be looked at, but then you'd have to take into account whether there were other drugs available for the same indication and so on.

So there's data out there. We

could look before and after. And for the ones that are upcoming, there may be already drugs on the market and maybe REMS are in the development stages for them so that these kinds of behaviors in prescribing and maybe even compliance could be looked at before and then after.

Concerning the patient access to the drugs, knowing what a healthcare provider says to the patient is helpful. That came up this morning in a simple question about, did your doctors give you any reservations about prescribing this drug, and then you can look in terms of what is in the REMS. Is it only one thing, you need to read a Medication Guide? That was already part of it. Or you have to have lab tests and so on.

So you could start parsing out what are the factors that contribute to perceived burden by the patient that might make him or her decline to go forward with the drug or not comply.

407

And so there are also simple questions that you can ask patients if you had this drug versus this drug, and there are the various criteria or requirements, how likely would you be to go forward with one of the drugs versus the others.

But again, do it in a hypothesis-testing way where you find out what are the components that drive it up or down so you have a better idea about what the burden really is, perceived by the patients, and then of course, perceived by the prescribers and whether they prescribe.

MODERATOR TOIGO: Thank you, Dr. Day. Dr. Gilsenan.

PANEL MEMBER GILSENAN: Just a couple of points with regard to assessing behavior at least of prescribers, you can use what Dr. Wolf was saying earlier, electronic health records or databases to look at whether or not drugs were prescribed to the appropriate populations. In some

circumstances some of the information you would like to have is not available or accessible in a population, a sample size that's large enough.

But I think that is a useful source for many things, and is important to look at. And then with regard to assessing patient burden, I think it is important to understand the people that don't get access to the drug, but you can also, it is important to ask the people who are taking the drug and assess the burden for, you know, what -- how they -- the perceived burden.

And one thing that may not be -- come to mind immediately about the burden but one thing that these REMS do is they communicate the risk over and over and over and over again to the patient, and there's no information about the benefit.

And that is very scary to patients, to see that information over and over again, and at some point, they just kind of -- and

409

we've had this feedback, talking to patients, they just put it away and they just don't even want to hear it anymore, they just -- they have decided they are going to take the drug and they just don't want to be reminded over and over again about the possible risks because they've gotten past that point. They have made the decision.

And so I think in some -- you know, that's a type of burden, and then you can ask the healthcare providers what the, you know, the burden to them is to administer the REMS.

MODERATOR TOIGO: Thank you. Dr. Davis.

DR. DAVIS:. These are just some things I was thinking. One is a response to that. One of the key messages of health literacy is patients have -- you need to figure out what they need to know, what they need to do and then why. It's always important to stress the benefit. Why is it that you want -- that you think this is in

their best interest.

But one of the things I just want to put on the table is, I'm intrigued with this sort of, this theme underneath this, is can this be educational as well as evaluation?

And so what you want from -- the goal it seems for the patient is safe use, and the goal for the provider is, besides appropriate prescribing, would be are they communicating, are they giving the patient education that's useful to them?

So one question I have for you all is, is that, does that have to be the MD? Can it be his or her staff? Can it be somebody else that's providing that? What's the pharmacist's role in here?

And then, just as a quick vignette, we did a study on birth control pills and public health years ago and the way we assessed these young women was we had Tanya questions, which were clinical vignettes.

411

Tanya goes away on the weekend with her boyfriend, forgets her birth control pill. She's gone for a week, so what -- you know, she misses one pill. What happens? She misses two? What happens?

And all the things we need to know, or about Tanya and the birth control pills that she -- the package that she had that was prescribed, with a calendar so it was all very concrete, and the patient said, the girls said they learned more from answering the Tanya questions than they did from 20 minutes of patient education they got from the public health nurses.

MODERATOR TOIGO: Thank you, Dr. Davis. And did, Mary or Rob, did you want to comment?

ASSOCIATE DIRECTOR WILLY: No, I'd just comment that it's an interesting question, that I think we could take back and talk about some more.

MODERATOR TOIGO: Dr. Holmes.

412

PANEL MEMBER HOLMES: I just want to ask a clarifying question of Drs. Shiffman and Moratto about how you capture patients and providers who decide up front they're not going to -- like, how do you do that?

Where's that sample, where's the sample frame? Is it all theoretical? Is it all modeling? Or is there actual approaches that you know and recommend?

PANEL MEMBER SHIFFMAN: Well, I don't think there's a well understood, good solution. One of the things I suggested is a clear indicator of a patient who may have been dissuaded because of the burden is someone who got a prescription and never filled it.

Now there could be other reasons and that would be, we'd be back to surveys basically, of that population and you'd want to understand why they didn't fill it. It could be lots of reasons, like their insurance didn't cover it.

But you could discern if the reason was that they then had to go to a pharmacy that was, you know, two miles away, and they had to fill out paperwork and felt uncomfortable, et cetera.

With providers, I think it's a little harder and the only way I have been able to think about doing this is to look at providers who, by the nature of their specialty and practice, would be, for your sales force, good candidates for that drug but who don't sign up, don't, you know, with a REMS that requires registration or showing that you have completed the educational materials, to survey some of the people who didn't sign up, are now not eligible to prescribe and find out whether it was because they weren't willing to put up with the burden.

I think one could also, it's not just, now, that's saying that that's targeting a provider who has taken him- or

herself out completely, but I think it would be reasonable to ask providers who did register, do you ever, you know, do you avoid prescribing this drug versus another because the burden to you or to the patient is too high.

And when we have asked questions like that, we get an earful. So I think people are eager to talk to you about it.

MODERATOR TOIGO: Dr. Morris.

PANEL MEMBER MORRIS: I'm really glad you are asking this question because it comes up a lot. I guess the thing that I want to suggest is that in addition to the actual burden, there's an important signal value of the risk management plan itself, in that you are communicating to doctors primarily that you know, something is really scary here about this drug, and there's a lot of refusal, especially among primary care, you know, specialists might be willing to take more risks because they need to, but

415

I've done enough focus groups with primary care doctors to know that when there's a REMS they get scared and they don't want to prescribe the drug.

So one of the burdens, I think, is really how much are we over-warning beyond what we should be warning in terms of communicating risk and it brings up -- that brings up another issue and that is, you know, to what extent can we know if we are over-warning or not, I mean, it's such a value-laden issue, and how do you know how much warning is enough, you know, and I think sometimes getting that perspective is incredibly difficult.

MODERATOR TOIGO: Dr. Moratto.

PANEL MEMBER MORATTO: I would agree, and just to kind of, towards the question of trying to quantify some of this, which you were asking, you can also do the thought experiment around what are the intended behaviors and what might be the

416

unintended behaviors that the REMS might induce, which could be people switching to other drugs or things like that, so you can use the same pharmacoepidemiology database to model both.

So for instance, if a drug says you should get this added test, did that happen; if they don't like that they might switch to other drugs if there's others in that class.

So in some way to say, you know, what are the other impacts of the REMS. And then to the point you had mentioned Dr. Shiffman, around the collective burden, you know, it made me think about, we can use these same data sets to look at different groups of doctors, let's say a family medicine doctor, you know how many total drugs are they commonly prescribing that now have a REMS and you could look at the claims databases this way, what are the number of patients that are affected by this and what's sort of the overlap of it, how many patients

417

do they have that are going to be affected by multiple REMS?

And that might be a way just to kind of quantify the overall population level burden of all of this, and to what degree, if you wanted to go into a content analysis, to what degree are they asking for similar things, is there a way to get more efficiency around that, to what degree are they asking for very different things, but it might be a way, as you are looking at, how many, 80-plus REMS, and where are they, as I was saying earlier, similar and dissimilar for similar kinds of risks in terms of the approaches.

MODERATOR TOIGO: Go ahead, Dr. Shiffman.

PANEL MEMBER SHIFFMAN: Just briefly, to comment on the issue of burden, that one has to consider the patient population involved so there are some opiate medications that are subject to REMS that are meant for example for breakthrough pain in

cancer.

So then you are dealing with a patient for whom dealing with someone else's goddamned REMS is just not a very high priority. So what would be a very minimal burden to a patient who, you know, maybe has a chronic condition but is generally healthy, can be a huge burden to a person who is dealing with cancer, may be dealing with end of life issues, and so the metric for how we estimate the costs of what may be the same behavior has to take into account the condition of the patient.

MODERATOR TOIGO: Thank you, Dr. Shiffman. Any additional comments on this question? Oh, Dr. Hornbuckle.

Well, we took them all together. We tried to manage throughout the day and that didn't work, so I just sort of put them out there and see where we go.

PANEL MEMBER HORNBUCKLE: For the second question, what are our alternative

419

methods to assess behavior, burden and access for a REMS. One thought is that we have heard these common themes throughout the day, that REMS assessment methods are not a one size fits all. I think we all would agree with that.

One approach is to really look at REMS assessments and REMS tools in a tiered approach, take a tiered approach and think about a pyramid, and that pyramid has the various REMS assessments and then the tools of what they are measuring.

And at the bottom of that pyramid would be the label, and then starting from the label, you would move up then to the Med Guide, the physician/prescriber information, and then the third level could be knowledge and utilization of surveys, with surveys there is also discussion of whether or not surveys could also measure behavior, whether or not can it also access or assess burden.

Then the next level would be

looking more at assessing behavior, which could be through secondary data analysis or use of secondary data. We have heard numerous panel members mentioning about the use of the pharmacovigilance post-marketing adverse event data, the electronic health records.

And then you look up at the top, which is really around behaviors around outcomes. All REMS really do not have to really look at behavior. Some of them are more communication. So it depends on the REMS goals and objectives on how you would require or have companies or -- to consider these assessment tools, because if you think about behavior and outcomes, one way to look at if a REMS has a serious risk, how -- of no fatalities, or minimizing fatalities, what would be an appropriate outcome to evaluate for that REMS?

Well, how many fatalities were reported or identified via that REMS? So if

421

you found zero fatalities, is that an effective REMS? Or if you found very few, is that still effective then, having a large number?

So behaviors as well as outcome would be more customized to the REMS goals and requirement per each medication.

MODERATOR TOIGO: Thank you, Dr. Hornbuckle. Any other -- okay, Dr. Eggers and Dr. Day.

PANEL MEMBER EGGERS: I'd just like to make a brief comment that the comment that I made with all the prefaces, those still apply, earlier, I think applies here as, I mean, maybe even more so as well, that -- and I will, I guess I'll more strongly promote the behavioral decision research paradigm which can incorporate FMEA, it can also incorporate things more like influencing diagram techniques to really understand the -- what's causing the root-cause burden problems or knowledge problems as a way --

422

and then it can couple, so I would think it could sort of address in one shot both knowledge issues and burden issues and as well, address benefit, sort of benefit-risk issues, sort of the flip side of burden, all in one shot, because at the end of the day they are all influences on decision-making and I know that that's not a direct -- that's not usually the stated goal of REMS, but the goals of REMS usually are all related to that -- helping to improve decision-making in some way.

MODERATOR TOIGO: Thank you Dr. Eggers. Dr. Day.

PANEL MEMBER DAY: I appreciate all the comments about one size does not fit all, and that's very true. It's also true that all REMS are there for a similar kind of reason, so it would be wonderful if we could determine some core features, whether they would be questions or whatever they are, that could exist across all of them, and it could

423

be, off the top of my head, something like a judgment of the people involved, whether the providers or the patients, overall, how risky do you think this drug is or how safe is it.

But some core questions that are implied by all of -- or involved in all kinds of REMS that would be constant across it all. So the idea of standard questions or elements versus those that are developed because of the different sizes are not going to fit all.

MODERATOR TOIGO: Thank you, Dr. Day. Dr. Holmes.

PANEL MEMBER HOLMES: I concur with that and -- but I still think, given your example, the one size doesn't fit all still holds, because let's say you had that question for all REMS, the threshold would not be the same for all REMS.

So no matter what, there is a one size does not fit all even if there's 5 or 10 questions that we ask in every REMS, if you follow what I mean.

424

PANEL MEMBER DAY: Correct. You could still ask it, though, and then you could contextualize it as a function of these other things.

MODERATOR TOIGO: Dr. Willy, do you have what you need on that question or shall we go one by one and see if anyone has comments on either of those questions? Did you get what -- is the discussion what you are looking for?

DEPUTY DIRECTOR SHIBUYA: We're okay.

MODERATOR TOIGO: You're okay? Okay. So then this is the point, Rob, where you would summarize what we heard from this afternoon's panel.

DEPUTY DIRECTOR SHIBUYA: I'd like to thank Panel 2 for, again, a very wide-ranging and thoughtful discussion. It was a good brainstorming session. I learned a lot of new terms. So it was useful for me.

And thank you for explaining some

425

of the terms that I didn't know. I'm going to be very quick about this because again, I personally have to do a fair amount of research after this meeting to understand exactly what was being proposed here.

But with regard to the first question about alternatives to knowledge surveys, we were given a good list of possibilities although I thought -- I think Dr. Morris or Dr. Shiffman fairly strongly stated that pretty much the only way to know whether or not somebody knows something, is to ask them, which in a way is a survey.

And the -- it sounded to me like the other methods that were proposed would sort of, in that, as Dr. Moratto was saying, form sort of a mosaic and sort of tangentially inform knowledge.

There were questions brought up about the feasibility of some of these methods. And I think that's about all I -- in a very broad -- painting with a very broad

brush, that's what I captured from the discussion of question 5.

With regard to question 6, I thought I heard concern about its -- the difficulty in capturing patients who have lost access to the drug. Again, we heard a lot of good ideas, how this could be possibly done by asking questions and also by using other databases available to us.

We heard a fair amount of interest in the systems engineering approach to looking at some of these issues. And then, not -- as Terry was saying, not directly related to today's discussion, but we did hear, and which we have heard from other public sessions, that FDA needs to work very hard to coordinate REMS to try to minimize burden and address the access issues, and I assure you, we are working very hard towards that end.

I think those of you who follow closely what we are doing are aware that we

427

are taking that charge from FDAAA very seriously.

That's about all I was able to capture. Terry, I don't know if you wanted to --

MODERATOR TOIGO: I just want to be sure that anyone who had a question from the audience was able to provide them to Julia. I thought there were more, we only had the one. We don't have an answer for it, and if you have any other questions, we may not have answers, but at least I want to afford you the opportunity to get the questions on the table.

Okay. So then I want to thank the industry representatives, our panel members from academia and consulting firms, and any -- all of the panel members, the FDA staff who have worked really hard in trying to put this workshop together, and I think it was a good forum for public discussion on a very challenging topic that we have been wrestling

with.

We learned that it is even more complex than we had thought when we came into the meeting. We have more questions. But I think we have some, some things that we can actually go back and sit and talk quite a bit about, that may be helping us in the future.

I remind you that, if you leave and you still have thoughts about things that would be important for us to know, the docket is open until July 7th. If you need help in figuring out how to submit comments to the docket, we can help you with that.

But we do really, we value those comments and we carefully consider them. And then lastly, feedback about the meeting, again, negative and positive, is helpful for us in future meeting planning.

There was a lot of discussion about how do you set up this type of meeting, do you have individual panels, small panels; we came up with this as final. It has its

429

pluses and minuses. But you participated, especially the panel members who participated in this, if you have thoughts to share with us on how you think it might have gone better, I -- you can call me or you can email me because I really do take those suggestions seriously.

So I think, with that, we can adjourn the meeting early and that's a good thing. So thank you, all, and safe travels.

(Whereupon, the above-entitled matter went off the record at 4:10 p.m.)