AT

DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

NONPRESCRIPTION DRUGS ADVISORY COMMITTEE

Monday, September 25, 2006

8:00 a.m.

Hilton Washington, D.C. North/Gaithersburg
Gaithersburg, Maryland

PAPER MILL REPORTING
Email: atoigo1@verizon.net
(301) 495-5831

PARTICIPANTS

Eric Brass, M.D., Ph.D., Acting Chair
LT Darrell Lyons, BSN, Executive Secretary

COMMITTEE MEMBERS:

Ernest B. Clyburn, M.D.
Jack E. Fincham, Ph.D.
Ruth M. Parker, M.D.
Robert E. Taylor, M.D., Ph.D.

CONSULTANTS (VOTING):

Neal Benowitz, MD.
Louis Cantilena, M.D., Ph.D.
Ralph B. D'Agostino, Ph.D.
Terry C. Davis, Ph.D.
Ruth S. Day, Ph.D.
Marie R. Griffin, M.D.
Richard A. Neill, M.D.
Wayne R. Snodgrass, M.D., Ph.D.

PATIENT REPRESENTATIVE (VOTING):

Musa J. Mayer, M.S.

INDUSTRY REPRESENTATIVE (NON-VOTING):

George S. Goldstein, M.D.

CONSULTANTS (NON-VOTING)

Saul Shiffman, Ph.D. Guest Speaker)
Alastair Wood, M.D. (Guest Speaker

FDA PARTICIPANTS:

Charles Ganley, M.D.
Andrea Leonard-Segal, M.D.

C O N T E N T S

Call to Order/Introduction of Committee
Eric Brass, M.D., Ph.D. 4

Conflict of Interest Statement
LT Darrell Lyons, BSN 7

Welcome/Opening Remarks
Andrea Leonard-Segal, M.D. 12

FDA Presentations

Health Literacy
Ruth Parker, M.D. 36
Terry Davis, Ph.D. 51

Consumer Behavior Studies
Saul Shiffman, Ph.D. 71

Information Processing
Ruth Day, Ph.D. 102

Statistical Considerations
Ralph D'Agostino, Ph.D. 126

Complexities of the Rx to OTC Switch
Alastair Wood, M.D. 162

Final Questions from the Committee
to the Speakers 193

Open Public Hearing

Dr. Sujit S. Sansgiry, University of Houston 213
Julie L. Aker, Concentrics Research 222
Douglas Ws. Bierer, Ph.D., CHPA 235

Committee Questions/Discussion 245

P R O C E E D I N G S

Call to Order

Introduction of Committee Members

DR. BRASS: I am Eric Brass from Harbor UCLA Medical Center. I am pleased to welcome everybody to the meeting of the Nonprescription Drugs Advisory Committee.

I would like to begin by just going around the table and allowing the committee members to introduce themselves. George, we can begin with you.

DR. GOLDSTEIN: My name is George Goldstein. I am a retired pediatrician and retired from the pharmaceutical industry after that, 17 years in the first case, and almost 30 in the other. I am the Industry Liaison Representative to this committee.

MS. MAYER: I am Musa Mayer. I am the Patient Representative at this meeting. I am a breast cancer advocate and 17-year survivor.

DR. CANTILENA: My name is Lou Cantilena, head of Clinical Pharmacology at the Uniformed

Services University, former member of NDAC in the early nineties for four years and then came back later on around 2000 or so and chaired the committee for a few years.

DR. NEILL: Hello. I am Richard Neill. I am a family physician at the University of Pennsylvania, a former NDAC member.

DR. SNODGRASS: I am Wayne Snodgrass. I am a pediatrician and head of Clinical Pharmacology at the University of Texas in Galveston, Texas.

DR. DAY: Good morning. I am Ruth Day. I am the Director of the Medical Cognition Laboratory at Duke University, a former member of the Drug Safety and Risk Management Advisory Committee, and served on this committee in 2002 last.

DR. DAVIS: I am Terry Davis. I am a Professor of Medicine and Pediatrics at LSU Medical Center in Shreveport, Louisiana. I am a psychologist and my area is health literacy. My missing colleagues will be here momentarily, Ruth Parker, internist at Emory, who also does research in health literacy.

LT LYONS: My name is Darrell Lyons. I am the Designated Federal Official for the committee.

DR. TAYLOR: I am Robert Taylor. I am an internist and clinical pharmacologist. I chair the Department of Pharmacology at Howard University. I am a member of NDAC.

DR. FINCHAM: My name is Jack Fincham. I a Professor of Pharmacy Practice and Public Health at the University of Georgia.

DR. CLYBURN: I am Ben Clyburn. I am in Internal Medicine at the Medical University of South Carolina.

DR. GRIFFIN: Marie Griffin. I am an internist and pharmacoepidemiologist. I am a Professor of Preventive Medicine and Medicine at Vanderbilt.

DR. D'AGOSTINO: Ralph D'Agostino, biostatistician, Boston University. I am also a former member of NDAC and at one time also chair.

DR. LEONARD-SEGAL: Andrea Leonard-Segal, Director of the Division of Nonprescription Clinical Evaluation for FDA, and Dr. Ganley, we

assume will be here shortly. He is the director of the office.

DR. BRASS: Darrell, if we could have the reading of the Conflict of Interest Statement, please.

Conflict of Interest Statement

LT LYONS: The following announcement addresses the issue of conflict of interest and is made a part of the record to preclude even the appearance of such at this meeting. This meeting is being held by the Center for Drug Evaluation and Research.

The Nonprescription Drug Advisory Committee will consider issues related to the analysis and interpretation of consumer behavior studies conducted to support marketing of nonprescription drug products.

Unlike issues before a committee in which a particular product is discussed, issues of broad applicability, such as the topic of today's meeting, involve many industrial sponsors and academic institutions. The committee members have

been screened for their financial interests as they may apply to the general topic at hand. Because general topics impact so many institutions, it is not practical to recite all potential conflicts of interest as they apply to each member.

The Food and Drug Administration has prepared general matters waivers for the following Special Government Employees:

Dr. Neal Benowitz, Eric Brass, Ralph D'Agostino, Terry Davis, and Marie Griffin, who are participating in today's meeting.

Waiver documents are available at the FDA Dockets web page. Specific instructions as to how to access the web page are available outside today's meeting room at the FDA information table. In addition, copies of all waivers can be obtained by submitting a written request to the Agency's Freedom of Information Office, Room 12A-30 of the Parklawn Building.

The FDA acknowledges that there may be potential conflicts of interest, but because of the general nature of the discussions before the

committee, these potential conflicts are mitigated.

Dr. Alastair Wood has been invited to this meeting on behalf of the agency. He was a past member. He is no longer SGE or consultant to the committee.

With respect to the FDA's invited industry representative, we would like to disclose that Dr. George Goldstein is participating in the meeting as a non-voting industry representative acting on behalf of regulated industry. Dr. Goldstein's role on this committee is to represent industry interests in general, and not any one particular company. Dr. Goldstein is a retired employee of Sterling Drug.

In the event that the discussions involve any other products or firms not already on the agenda for which FDA participants have a financial interest, the participants' involvement and their exclusion will be noted for the record.

With respect to all other participants, we ask in the interest of fairness that they address any current or previous financial involvement with

any firm whose product they may wish to comment upon.

Thank you.

DR. BRASS: Thank you.

The topic of today's meeting will focus on the design, analysis, and interpretation of consumer behavior studies used to support OTC switch applications.

Having been on this committee for some time, there is a tendency to look at meetings that revolve around general issues like this as being of somewhat secondary importance to NDA meetings or issues that are more focused simply because it is hard sometimes to get a tangible output out of meetings like this.

But I want to assure everybody that is participating in today's meeting and those who are affected by the discussion that this issue today is going to be of extreme importance going forward in how this committee and the agency looks at these types of trials, and the issues that we are discussing are of substantial importance to both

the industry regulators and to the public health, so I thank everybody in advance for their participation.

The discussion today is going to be extremely interesting and important, and I am going to try to preserve time to focus on that discussion, so I thank all speakers in advance for complying with the time, which is designed to ensure that the committee members have an opportunity to discuss the issues.

I would also like to remind people that the studies we are talking about have as their objective to model and predict consumer behavior in the real marketplace, and I think that is an extremely important concept to keep in mind.

These are not academic exercises, these are not designed in vacua, these are not designed to meet some kind of arbitrary benchmark, but the true benchmark is whether or not they allow us, as a committee, and the regulators to make predictions that are accurate as to what will happen if a specific drug is made available in the general

marketplace OTC.

I would suggest keeping that overarching objective in mind is helpful in providing perspective.

With those brief comments, I would like to turn the floor over to Dr. Segal to provide some opening comments. She has already told me she would go overtime, and I have said that's okay.

Welcome and Opening Comments

DR. LEONARD-SEGAL: First, I want to thank Dr. Brass for being so flexible with me this morning. My talk will probably go about two or three minutes over, but, in fact, we are starting a little early and that's great, because I have been given the pleasure of presenting some awards that I didn't know that I was going to be presenting this morning, so I think that we will do that first and then we will get on to the talk.

First, Dr. Wood. These are awards to thank committee members that are with us for the last time on the committee this morning.

Dr. Wood served as chairman, and so ably

as chairman, and it is a pleasure for me to give you this award this morning. Dr. Wood will be speaking to us today, as well.

[Applause.]

I guess I can read that it says, "In recognition of distinguished service to the people of the United States of America."

[Applause.]

DR. LEONARD-SEGAL: Dr. Benowitz. Dr. Benowitz has served as a very able committee member for the last four years I guess, right? And your award says the same thing, so the people of America thank you, too.

DR. BENOWITZ: Thank you very much.

[Applause.]

DR. LEONARD-SEGAL: Dr. Snodgrass, who has also served for the last four years, and your award also says the same thing. Thank you very much.

[Applause.]

DR. LEONARD-SEGAL: Dr. Brass, members of the committee, invited speakers, it is a pleasure for me to welcome you all this morning. I have got

to say that we are very enthusiastic about today's meeting, because it is going to give us an opportunity we think to advance the science of over-the-counter drug development.

[Slide.]

Let me remind you that there are three types of studies that are conducted to predict consumer behavior with other-the-counter drugs. These are label comp studies, self-selection studies, and actual use studies.

This morning I am going to raise lots and lots of questions about trial design issues and analysis for these three studies, but as a backdrop, I think it is really important to emphasize that we think these studies are predictors of consumer behavior in the over-the-counter setting, but they have not been validated.

[Slide.]

Unlike results from randomized controlled studies where a drug can fail to demonstrate efficacy and/or safety, we think that there

probably are no failed consumer studies, because we can learn something about consumer behavior and labeling from each one of them, and hopefully, each one can lead us to conduct a better subsequent study.

[Slide.]

Now, over the next 20 minutes of so, I am going to be talking about labeling, the three types of studies. I will give you the charge for this morning and the agenda.

[Slide.]

First, labeling. It is important to note that information necessary for correct self-selection must be on the Drug Facts label. That is an imperative of the Drug Facts label.

Lately, we have seen over-the-counter products and proposed products for which the labeling is more and more complex as with the cholesterol-lowering drugs that have come before this committee, and the over-the-counter NSAIDs, which have been out there for quite a long time, but now sport new organ-specific warnings which

really populate the label to a magnificently cramped extent.

[Slide.]

So, I ask you at what point do we pack so much information into the label that people stop reading it, how should we determine what information must go on the Drug Facts label for self-selection, and what could go into a package insert.

Let me just say that information on inserts, sometimes called consumer leaflets, can be a condition of approval and are labeling subject to FDA regulation. We do control what is on the package insert.

[Slide.]

We have a lot of products with package inserts that are currently marketed over the counter - Today Sponge, the vaginal anti-fungals, Plan B, a new one, and these inserts tend to contain expanded information about directions of use.

Sometimes they provide pictures for the

consumer to refer to, so that they can hopefully use the product more precisely, and they will also even sometimes provide information about the condition for which the product is intended.

Dr. Day will talk with us this morning about information processing, and she will address our labeling I believe in that context.

[Slide.]

Let's move to label comprehension studies, why do we do them and what do we want to know more about.

[Slide.]

I think we could say that the purpose of label comprehension studies as they have evolved is really twofold. One is to test how the label communicates information to the consumer, and the other would be the ability to test the ability of the consumer to apply label information in hypothetical settings in which the drug should or should not be used.

[Slide.]

It is important to remember, though, that

understanding words does not necessarily predict decisions and actions, and one thing that we have seen is that good label comprehension study results do not necessarily predict good results in actual use studies. However, we have also seen that poor label comprehension study results may predict poor results in actual use.

[Slide.]

So, I ask you are there good ways to improve the correlation between label comp results and actual use results bearing in mind that thus far the two types of studies generally have enrolled different populations, the label comprehension studies enrolling all comer type populations, and the actual use studies tending to enroll interested users.

[Slide.]

Now, let's turn to literacy. We are going to hear a lot about health literacy this morning from Drs. Davis and Parker. OTC labels have been targeted to an 8th grade literacy level, and the populations enrolled in the label comprehension

studies have consisted of both normal and low literate participants, and we have called that group the general population.

In addition, we have tended to enrich these studies with more low literate participants, in other words, those that read at less than an 8th grade literacy level.

Also, note we get all this information and it has not been clear how to use it, so how should we use low literacy information that we obtain from these studies, should the normal and low literacy populations be analyzed separately, or should they be analyzed en masse as one general population.

[Slide.]

Does comprehension need to be the same for the normal literate and low literate populations, and, if not, what degree of difference should we tolerate?

If 90 percent of the normal literate population, for example, understands that a person with kidney disease should not take a drug, but only 70 percent of the low literacy population

understands this, how should we act on this information?

[Slide.]

What is it reasonable for us to expect a consumer to be able to understand from a Drug Facts label? Often, decisions about communication success have come down to whether the comprehension level feels good enough to those interpreting the data.

So, I ask you, do we expect too much of consumers, or do we not expect enough of them, and how do we determine what is adequate comprehension for a particular label communication element, how do we know when to stop testing the label, how do we know when we have achieved as much as we can?

[Slide.]

Now, during the label comprehension study, the study participant is usually handed a copy of the label, which they can study as long as they want to, and then with the label in hand, they answer the questions that they are asked by the questioner.

Now, this is not naturalistic, because I don't think that anyone goes into a drugstore and studies the package label for half an hour or 20 minutes at a shot.

Does this methodology, therefore, inflate the comprehension results, could the methodology be improved?

It is also important to remember, though, that label comprehension testing that might require the participant to remember what is on the label, for example, letting them read it and then taking it away, and then questioning them would also not be naturalistic, because it means that they have to memorize what they have seen.

[Slide.]

Commonly, industry will ask us this question: Are there answers that are not precisely correct as per the label information that could be considered to be acceptable?

So, I ask you, is comprehension black and white, or is there a gray zone?

Should there be acceptable label

comprehension study responses and how do we determine what those would be? Industry has often grouped acceptable answers with correct ones. How should we analyze correct answers? I hope I am getting your wheels turning here.

[Slide.]

So, consider this scenario. The label warning says, "Stop use and ask a doctor if you have abdominal pain." The scenario is Sam is taking Drug X. He develops abdominal pain. What should he do?

The correct answer is stop use and ask a doctor, but the respondent answers, "Ask a doctor." Now this is not precisely correct. One could say it's even a default answer, but it could be acceptable. How should we interpret answers like this, is this correct label comprehension, is it acceptable label comprehension?

[Slide.]

Let's turn to sample size. Industry often asks what an appropriate sample size is for the general population, as well as for subpopulations.

Generally, we have seen about 300 normal literacy participants in these studies and about 150 low literacy participants. It is unclear that these studies are always sized appropriately and we would like better clarity as to how to populate these studies. Dr. D'Agostino is going to talk to us about statistical issues and sample size with regard to all three studies that we are considering this morning.

[Slide.]

Let's move on to self-selection. What are they, why do we do them, and what do we want you to be thinking about this morning?

[Slide.]

The purpose of a self-selection study is to determine if the consumer can correctly decide whether or not the product is appropriate for him or her to use based on the label information. These studies may be stand-alone studies, we have seen them that way, or they can be tagged on as part of a label comprehension study or as part of an actual use study.

[Slide.]

One of the burning questions is how to ask the self-selection question. A typical type self-selection question might be is it appropriate for you to use this product. It is not clear if we are asking the question the best way to acquire the information that we want to know, so I ask you this morning to be thinking about what is the best way to ask the self-selection questions so as not to influence the respondent.

[Slide.]

Now, say that we have got a self-selection study. This is what we have done. The participants make a self-selection decision and some say "no," this product is not right for me to use.

We have tended to disregard those people and we have put all of our emphasis on the ones that have answered "yes," and some of those have answered correctly as per the label information and their own history, and we have accepted those answers.

There have been others that have answered

incorrectly, and we have wanted to know why they have done that, so that we could learn more about what they were thinking.

[Slide.]

Let's go back and think a little bit more about this "no" group that sort of falls off our radar screen. Should we continue to disregard them, or should we only care about those who say "yes," because they are the ones that are going to be taking the drug, and that is sort of the way our thinking has been running, but the "no" self-selectors could be correct self-selectors, so consider this scenario.

We have 1,000 people who make a self-selection decision, and 900 of them answer "no," and it is possible that they are all correct, the product really might not be appropriate for them, or maybe even taking it one step farther, some of them didn't understand the label and they are going to say "no," I don't want to use this product because I don't want to take something I don't understand.

One hundred of them say "yes," and half of them are correct and half of them are incorrect. Now, if we look at the "no" group, 900 plus the 50 correct, that would be 95 percent correct self-selection. That is pretty great.

On the other hand, if we just look at the correct group of the "yes" answerers, 50 percent have answered correctly. That is 50 percent correct self-selection. That could be the difference between a successful or a failed study depending on interpretation.

[Slide.]

Now, let's go to the "yes" responders, and I ask you, when is incorrect, in fact, acceptable. Incorrect self-selection decisions to use a product for one person may be acceptable for another based on the individual's unique medical history.

Consider cholesterol-lowering drug. Women over 55 years of age can take the drug. They would be correct self-selectors if they are in that age category. A 40-year-old woman self-selects to use

the drug. That is incorrect.

A 40-year-old woman that we know something about, we have got some Y answers. She is status posthysterectomy, and her mother died of an MI at the age of 36. Her answer, her decision to use this product might be considered acceptable. Should this acceptable answer then be analyzed as correct?

All I can say is that it appears to be important to collect information about why consumers make self-selection "errors" in quotes. Often, sponsors have not been providing us with that information.

[Slide.]

How should we interpret self-selection data? For a product label that is comprised of an indication with multiple components and multiple warnings, do participants need to weigh every piece of information correctly in their decision-making?

Now, this is something I believe Dr. Wood is going to be talking about in his presentation among other things. So consider the cholesterol-lowering population. There is an awful

lot to be thinking about to make a correct self-selection decision.

[Slide.]

What we have done, what we did do with one of the self-selection studies that we saw in cholesterol-lowering drugs was we looked at the percent of perfect responders and found that there were only 5 percent of them.

[Slide.]

Was this too stringent an approach on our part? Could we have used different types of self-selection decision analyses?

For example, cumulative scoring of the different elements, so that people might need to get maybe 5 out of 6 of them correct, or could we have predefined a hierarchy of importance of the different elements, and if we were going to do that, how would we prioritize what those elements might be?

[Slide.]

Once people make a self-selection decision, should we verify it? This can be a very

difficult thing to do, and we wonder how aggressively we should pursue this. Now, for cholesterol-lowering drugs, we did request verification in the form of lab data.

But I ask again, is self-reported information from study participants sufficient, would it be sufficient for laboratory tests, would it be sufficient for other historical information? Do we need to confirm that a study participant spoke to a doctor? This can be a very hard thing to do for many, many different reasons.

[Slide.]

Now, let's turn to sample size in self-selection studies. It has been quite variable. The general population has often been tied to the sample size of label comprehension or actual use studies, and so we have seen self-selection populations that have ranged from a few hundred to thousands. Subpopulation samples size is also variable.

Recently, we reviewed a study looking at teenage self-selection that enrolled 150

participants, and we have recently reviewed two drug-drug interaction self-selection studies that each enrolled about 50.

How should we determine the size of the general population and subpopulation?

[Slide.]

Let's turn to actual use studies. Dr. Shiffman is going to be talking to us today I think about the self-selection studies and the actual use studies.

So, why do we do actual use studies and what do I want you to think about this morning? Actual use studies simulate the over-the-counter use of a product, and they can assess a lot of different things, for example, the relationship between a self-selection decision and a purchase decision, adherence, safety, and rarely they have been used to look at efficacy in the OTC setting.

[Slide.]

Often, these actual use studies have been single-arm, multi-center, uncontrolled, open-label studies, and we wonder if we should be considering

different designs, for example, multi-arm studies where we could evaluate different labels in actual use or where we could evaluate the benefit or lack thereof, of educational materials that are being proposed.

[Slide.]

What about that purchase decision? After making a self-selection decision, the consumer must decide whether to purchase the drug.

Sponsors often ask us to consider data on purchase decisions in actual use, but we have been uncertain as to whether this is a good idea, because we know that price influences purchase decisions and we cannot control the variability of drug costs.

Therefore, what is the relevance of considering the purchase decision of study participants?

[Slide.]

How long should actual use studies go on? This is another question that comes to us frequently, and generally for a short-term use

over-the-counter drugs, for example, an analgesic that is labeled for use for 10 days, we might have asked for studies to go on a week or two longer than the labeled duration of use to see if people stopped using the drug, to see if they start reusing it.

Is this appropriate? And for a chronic use drug, how long should we determine the appropriate study duration?

[Slide.]

What about adherence?

[Slide.]

We don't know what happens with prescription medication, although the assumption is the prescription world is generally a world of ideal use. But we know that patients are often noncompliant and we know that doctors sometimes prescribe the wrong drug, they make a selection error, and we don't want to set an unrealistic OTC standard for adherence.

So, the question is how should we determine what our threshold should be for overuse

or under use and adherence for chronic use products.

[Slide.]

Now, there are a few issues that all these studies have in common, so I will group them together now.

[Slide.]

First, population differences. Should thresholds for success and failure for label comprehension studies, self-selection, and actual use be the same across populations, and if not, how should we determine what the difference can be?

When should the majority who could benefit from access to an OTC drug be denied that access because of self-selection errors made by a subpopulation at risk from drug use? I think Dr. Wood will talk a little bit about this today, as well.

[Slide.]

For analysis, the results for the general population and subpopulations have generally been analyzed to determine the percentage of correct

responses for communication objectives in label comp studies, for self-selection decisions, and for the different actual use elements.

[Slide.]

But consideration needs to be given as to whether this data should be presented other than as a point estimate, perhaps as confidence intervals, and how these studies should be powered in the sample size calculated. I think Dr. D'Agostino will address some of this for us.

[Slide.]

So, your charge for today is to generate new ideas for better consumer research for over-the-counter drugs.

[Slide.]

This an overview of today's meeting. We will have our invited speakers this morning followed by questions from the committee. Then, there will be lunch, the open public hearing, and then I hope a very interesting and exciting committee discussion.

Thank you for what you are going to do

with us this morning, help us open up some of these issues for discussion and debate, and hopefully, forge a new and better path for our drug development process.

Thank you very much.

DR. BRASS: Thank you.

I would just like to make an observation for you to comment on. There is a tendency by some to think that what is going to come out of this meeting or what has been said before is a definitive template for every LCS, SSS, or AUS.

It seems to me as you are posited a number of questions, that, in fact, the answers are not singular and unique, but are highly dependent on the individual application, the individual drug candidate, and that at best what we are developing are guidelines to apply to unique situations to develop unique answers.

Would you comment on that?

DR. LEONARD-SEGAL: I think that you have it just right and we are not looking for specific answers today. I have thrown out a lot of

questions, and you will see that the questions that you are going to focus on later this afternoon are actually not designed to be looking for specific finite answers, but are really to open discussion and to give us different kinds of possible paths that we can think about and can try out.

We are hoping that we are going to have a lot of new ideas that come from today's meeting.

DR. BRASS: Thank you.

DR. LEONARD-SEGAL: Thank you.

DR. BRASS: Our next speakers apparently are a tag team of Drs. Parker and Davis, who have been allotted 25 minutes for their dual presentations.

FDA Presentations

Health Literacy

DR. PARKER: Thank you.

One of the most useful things I just found out, which my four teenagers would verify, is that I am not normal. I do go in drugstores and look at over-the-counter labels for 30 minutes at a time, I do it a lot, and I do it because of the interest

that I have in health literacy, so this will make my family members feel a lot better, they tell me this all the time, these teenagers, they tell me a lot more.

Anyway, thank you for the time.

[Slide.]

Terry Davis and I have been working together, and independently, as well, over the last 15, 17 years, to think about how well people can access and understand the information that they need in order to take care of their health.

We work to help define what it is we mean about health literacy and begin to measure its prevalence, to look at some of the associations, to spend a lot of time more recently advocating for improved health literacy in the country, and the good or the bad news for you all is that some have dubbed us "the Thelma and Louise of health literacy." That may be your fair warning before we take off here.

[Slide.]

Let me give you just a brief overview of

what we are going to try to do here. I am going to spend a little time at the very beginning here just defining what it is we mean by health literacy and what we know a little bit about how it relates to medication labels.

I am also going to use some real pictures of labels and just make sure we all are thinking about the same things when we say medication label. Terry is then going to talk a little bit about what we know about health literacy and medication labels from some studies we have been involved with over the last few years.

We are going to intersperse this conversation over the next few minutes with a few video clips of real patients, real people as they take labels and they take other critical health information and read it and tell us what it is they understand about it.

Most people remember that a lot more than they do anything that I ever say or Terry ever says, because I think what resonates is these are real people, we know these people. The scary thing

is you can't tell by looking who they are. So, you will see a few clips of some of our real patients that we have been working with and really learning from over the last few years.

We will end up concluding with some of our own thoughts about how to improve over-the-counter medication labels.

I think everybody here would really agree that labels are necessary for drug safety. I don't think there is much debate about that, and I think most of us would probably also agree that they are probably best considered a system of essential information for safe and effective use, information that is often for the prescribers, the pharmacists, certainly for consumers, and for OTC, the FDA has the authority to ensure the label communicates that essential information in a manner--and this is a direct quote--"likely to be read and understood by the ordinary individual under customary conditions of purchase and use."

So, FDA has the authority to ensure the label communicates essential information in a

manner "likely to be read and understood by the ordinary individual under customary conditions of purchase and use," all of which sounds pretty reasonable.

For over-the-counter, labels communicate information directly to consumers, and the FDA has that authority to ensure essential information is understood by ordinary individuals.

This is where the story gets really interesting, because the work that we have been doing in health literacy is really about what real people, ordinary people are able to access, understand, and use when it comes to real information, essential health information like that on labels. The bottom line there is not really good news.

[Slide.]

I worked with the IOM report on that committee on Health Literacy Prescription to End Confusion, and what we found were that 90 million adults have trouble understanding and acting on health information.

The definition that we used in health literacy there is the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate decisions.

Very importantly, the report also emphasized the sort of dual nature of health literacy. It is not just the skills that people come into it with, but it's the complexity of the task on the other end that they are asked to use their literacy skills with in order to be able to make and act on decisions.

So, real critically, this meant that literacy was not just can you pick it up and read it, but can you read it and understand it in order to be able to use it, and that is what is meant by the functional nature of literacy. It actually has to do with the ability to use information, which comes from taking it in, processing it, understanding it, but with the thing on the other end of actual use.

The report commented that complex tests

must be simplified and attention paid to culture and language, as well.

Well, in the last few weeks, some of you have probably heard about a very important national assessment of adult literacy. It came out of the Department of Education. It's the best portrait our country has of literacy skills in America.

Very recently, the report relating to health literacy skills, which is really the first national assessment of health literacy in our country, this is a survey of over 19,000 American adults. The results of that became available, so that we really now have a report of the health literacy skills of American adults.

[Slide.]

The headlines from that were really that 53 percent of American adults have intermediate health literacy skills. You will see the intermediate there, the biggest part of the pie, and that really is what captured the headlines in the reports of this.

Let's stop for a minute and say what

exactly does that mean. It looks like, well, you know, it could be worse, proficients up at the top, but what is that, what is the meaning of saying that 53 percent of Americans have intermediate health literacy skill.

Well, I will give you a hint. Take a look at the little arrow down at the bottom. The average American adult scored down lower, not in intermediate, but in basic. What were the tasks that were involved in this, because I told you literacy was not about just reading, it's about taking materials that relate to real tasks and being able to use the information that you read.

Well, for these 14 percent that scored in below basic, this meant that the people there had a 67 percent probability of accurately and consistently being able to perform a task like circling the date on a medical appointment on a hospital appointment slip, so 14 percent had a 67 percent probability of being able to accurately and consistently do a task like that.

Twenty-two percent scored at the basic

level. That means with 67 percent probability of accuracy and consistency, they could look at a clearly written pamphlet and give two reasons someone with no symptoms of a disease should still be screened.

All right. The average American scored in that basic range. Now, 53 percent scored in the intermediate range, with 67 percent probability these folks could determine what time to take a medication, a prescription medication based on the label on a prescription bottle that related the time of taking the medication to eating, and they could also identify 3 substances that may interact with an over-the-counter drug to cause a side effect using information on the over-the-counter label. Note the average American scored below that.

[Slide.]

Seventy seven million adults scored in the basic or the below basic health literacy range. These adults could not understand how to take a prescription medicine based on a common label

instruction, and they cannot identify 3 substances that interact with an over-the-counter drug based on the label. That is the best portrait that we currently have of the national average and of adult literacy in our country.

[Slide.]

Health literacy problems are common and they are probably here to stay.

Many of you may also have seen headlines over the last several months about high school dropout rates in our country, which on average are about 30 percent. I think my city can boast a rate significantly higher than that one.

I would like to turn to the video clip now and let some of our patients, real patients, real consumers show you some real tasks and their approach to common health literacy scenarios.

[Video played.]

[Slide.]

DR. PARKER: Seventeen, in case you were counting, Ms. Karr and her kidney transplant. I think the video just demonstrates so clearly a

couple of really important points. You really cannot tell by looking.

You also can't tell by asking the question do you have any questions. People don't tend to want to tell you no, I don't. I don't know everything I need to know in order to be able to take care of myself, it's embarrassing. You know, it's much easier to say no questions, I got it, I got it, and just sort of walk out not knowing, walk away not knowing, because it actually is embarrassing. So, you can't tell by looking. The tasks are more complex than you think, and there are a lot of them.

[Slide.]

I want to talk for just a few minutes very specifically about medication labels. For me, I always have to start at the beginning. I have got to go back, I have got to make sure I know what I am talking about with this, and I think the best way to think about medication labels is to think of this information as a system of information.

It is information, broadly speaking, for

prescribers, for pharmacists, for consumers, for patients, and it has many components potentially, some are more relevant to OTCs than others.

When we think of labels, you can think about the primary container labels. Now, for prescription medicines, this is State regulated, it does have FDA requirements. There are auxiliary labels, you know, those little stick-on warning labels that are different colors with little icons on them, industry-generated, not standardized, and there are these consumer medication information, CMIs that have little evidence actually to support what's on them, the quality of information, the content.

There are also medication guides that are a part of this system, prescription only. These are actually targeting providers. There are 28 of those now developed by the FDA. Those have been around since 1995, differing numbers of them.

Then, there are the PPIs. I love this one, Patient Package Information. Those are actually, when you talk about prescription

medications, those are specifically for the provider, yet, they are called PPIs. They are written by the manufacturer, they are approved by the FDA.

These are the different components for OTCs. The primary container label, which is not State regulated for the OTC, and then you have got the PPI that Laura mentioned, that are available for some OTCs.

[Slide.]

All right. A few pictures here. In the center you have got the PPI, you have got the drug up top, you have got the primary container label down there, you have got some stick-on warning labels.

[Slide.]

Here is another one. Here is a PPI for a prescription inhaler. Interestingly, this one was actually available in English and Spanish. I didn't find another one that had both languages, but that is an interesting point, why is that, why some, why not others, what are we doing about

language.

[Slide.]

Then, you have got these patient information leaflets. These are the things that are often stapled to the bags. Another piece, industry generated, that are a part of this potential system of information.

[Slide.]

Laura has gone over this. We know about the Drug Facts label, required standard format for all the OTC labels to give this easy-to-find information.

[Slide.]

This is from the FDA web site with the components that include the active ingredients, the uses, the warnings, the purpose, the directions, this other information. I think most of you are probably pretty familiar with what they look like. They are in your medication cabinet at home, you see them in the drugstore.

[Slide.]

Now, here are some--I actually made a

recent trip just to go check them out and see how they have changed lately--here are some that I picked up. I am going to invite you to come up and check out my bag later.

These are interesting and I want you to come up later during the break or whatever and take a look at some of the PPIs for some of these. These happen to be one of the categories that do require a PPI.

[Slide.]

As Laura noted, vaginal sponges, vaginal anti-fungals, H2 blockers, nicotine replacements, these are part of the H2 blocker PPIs. The proton pump inhibitors and the H2 blocks and their PPIs, and I just want you to take a look.

[Slide.]

Laura mentioned the kind of stuff that is on there, but let me tell you, in my trip to the drugstore, in buying these, one of the components that I found in two out of three of the ones that I purchased were coupons for buying more, and some of the coupons were actually double-sided where you

could get $2.00 off for coming back, so I have got some, I have got a bag of them actually, and I would like for anybody who wants to take a look and sort of think about them and decode what is really on these, what are these really doing, what do they really mean, take a look.

Terry has got another bag. So, I am going to turn now and let Terry take over and give a little information about some specific studies.

[Slide.]

DR. DAVIS: Here are some hidden problems with medication labels. There are tons of labels. The recent health literacy survey that Ruth was referring to found that over a third of U.S. adults cannot understand common dosing instructions on a pill bottle, poor understanding of instructions is a source of medication error and an issue of safety.

One of the points that we are going to make is that label instructions are simple, but that doesn't mean they are clear.

This is a clip showing you mistakes that

people of all literacy levels can make on over-the-counter prescription drugs and misunderstanding warning labels.

[Video played.]

DR. DAVIS: So, you know these people. Not only are they your patients, they are your neighbors and your family members. The other thing is that they all got a little bit right, but maybe they got just enough right to be dangerous.

[Slide.]

So, one of the points is that label instructions are simple, they are not clear, and they are not consistent. These are Ruth's patients. She asked patients to--this is the same instruction--1 capsule twice daily. Another one read 1 tablet by mouth twice a day for 3 days, 1 tablet 2 times a day, 1 tablet by mouth twice a day. This one was in Spanish. This one you couldn't read how many times a day, and the last one said take as directed.

So, if every time you get it, it's a different way to present it, that can be a little

confusing.

[Slide.]

We did two studies that focused on prescription meds, but I think apply to OTC, so that is why we are including them. We asked patients, "How would you take this medicine?" Forty-six percent did not understand instructions on at least one of the five labels; 38 percent with adequate literacy, above 9th grade, missed at least one label.

So, health literacy is not just a literacy problem, it is the connection between the skills you bring and the complexity of the task.

[Slide.]

Now, the other thing one of the women in the video pointed out is sometimes people can decode or read a label, but does that mean they can understand and use it. So, we asked them, "Show me how many pills a day you would take with these instructions: Take 2 tablets by mouth twice daily."

[Slide.]

If you notice the people with low literacy, in this case 6th grade and below, 71 percent could read or decode that label, but only 35 percent could tell us to take 4 pills. Now, this came, this research question came from the internists that worked with me, and they would come back and tell me stories just like this, so we tested it out.

[Slide.]

Then, we also looked at warning labels, and we asked patients, "What does this warning label mean?"

[Slide.]

Simple instructions, 84 percent got it.

[Slide.]

More complex instructions, about half. I am not an M.D., what is "plenty of water," how much water should I take?

[Slide.]

Then, this instruction, I love this, "Do not take dairy products, antacid, or iron preparations within one hour of this medication."

That is pretty complex, and I don't know what an iron preparation is. Why don't they just say iron?

[Slide.]

This table shows that the lower the literacy, the less likely you are to understand it, and the more complex label, regardless of your literacy, the less likely you are to understand it.

[Slide.]

So, if you can't read very well, is this helpful to you? If you can't read, what warning, what are you warned of? Here is what our patients said. "Someone swallowed a nickel." "Indigestion." "Bladder." "Looks like Casper, the friendly ghost."

[Slide.]

Does adding words help? Here is what some of the patient answers were. "Chew pill and crush it before swallowing." They didn't get the "do not."

"Chew it up so it will dissolve, don't swallow." "Just for your stomach."

[Slide.]

Now, what does this mean? This is my favorite. "Somebody is dizzy." "Someone is having an experience with God."

[Slide.]

Does this help, "For external use only." We use that a lot. "Use extreme caution." This is a common mistake with people with low literacy, kind of going for partial credit. "External" kind of looks like "extreme."

[Slide.]

So, my point is labels are short and simple, but they may not be clear. "Take one tablet four times a day," is that every four or six hours? What time should I take them? Do I take them with food, and what if I can't afford the whole bottle?

[Slide.]

So, the conclusion is health literacy is the skills plus the task complexity. U.S. skill level is not improving. U.S. dropout rate is very high. Ordinary adults lack adequate skills, and the task complexity needs improvement.

The research findings: Simple does not equal clear, current warnings are confusing, limited evidence for best practices.

So, the best way we think to improve OTC labels are to find a way to say it clearly, standardize it, don't say more than you need to. We need to figure out something about language and about icons and about advertising.

Thanks.

DR. BRASS: Thank you.

We have time for questions from the panel for Drs. Parker and Davis.

You have demonstrated convincingly many of the challenges consumers and patients face when confronted with medication labels. I am interested in extending that into what they do when confronted with health problems and those same labels.

I think here it may be different between a prescription and OTC drug. When they are given a prescription, they know they are supposed to do something with it, and therefore, they are likely to take it.

But if a consumer goes into a drugstore and sees a label they don't understand, will they purchase it and misuse it, or simply not purchase it?

DR. PARKER: There may be data on that. I don't know if there is, and if there is, it is not widely publicized and known. I can only tell you that the work we have done qualitatively. Terry has done hundreds of focus groups with people around the country about different issues related to health literacy, and in it, issues related to medication use commonly come up.

We do hear patients, and we have a couple themes that have emerged regarding particularly in areas, underserved populations, where we ask patients--I don't know if we published this or not, it may be in one of our papers, but we have asked them about when they have limited resources, and they approach the OTC setting to try to find something to reduce fever, they all know the marketing jingles, they have got the tag lines down. You can pop out a tag line, and people can

just tell you exactly what the product is.

We also have found a lot of patients who tell us that they buy name brand products, because they feel like it shows more love for their child or for their family member, and that is concerning, and efforts in one on one to try to educate people about that, there are groups of providers now that take this to heart, find it a very difficult up-hill battle to really explain limited resources, the use of those resources, and efficacy.

The safe use is difficult particularly when you take tasks like fever reduction, and you try to use two different products, and you try to explain how to alternate them, and you draw clocks, and you show them where to do it, and then you ask them in a setting clinically to say can you now review with me what I have just gone over to make sure I was clear in what I said, the tasks are really complex, and I think that is where the call for best practices is sorely needed.

I don't think it is impossible, but I think in terms of saying we know from evidence this

works and how to do it, we are still lacking on some of that, but you point out a very good problem, and I can only answer it qualitatively.

DR. BRASS: One other question. Do you have any data or any experience that relates to the comfort level, or knowledge or insight of low literacy consumers and the opportunity to interface with a pharmacist when buying OTC medications?

DR. DAVIS: Just qualitatively, and they report to us that they will ask their pharmacist questions. They find the pharmacists more accessible. Now, life is changing, as you all know, and they may be less accessible than they used to be.

DR. D'AGOSTINO: Your presentations were fascinating and I wonder if we should have any OTC product given the confusion you can have. But the question that I would like to ask, you end the presentations with the list of suggestions.

Are there data on the validity of those suggestions? I mean on the face they are certainly valid, but do we have a lot of literature to say

that in OTC or Rx settings, these things really do make a difference?

DR. PARKER: Let's go through the list. If we could just take a moment and walk through them.

No. 1 is find the best way to say it, which is clarity, and I would say that the data in support of that really comes from data about the confusion. I would say most of this goes along with more face validity than with actual outcome data, which is really your question I think, and it's a good one.

But I would also point out to you an every-day example in our life I think we can all understand and one I resort to a lot. I think if everybody in the room saw a stop light, and they were driving their car, and it was red, they would stop or they would know they should stop.

Now, not everybody would stop, but they would know they should stop. All the stop lights in our country are red, yellow, and green. They are not purple, some blue, some orange. There is a

standardized use of color.

There is a standardized format for what a stop light looks like, and if you had never seen a stop light, as people in other cultures haven't, you wouldn't intuitively know that it was a stop light and what it meant to do in traffic.

However, with standardized use, people can learn what it means, and it can become a part of what we do, and I think that concept of a standardized use allows us to educate people, so, in other words, the point about the clarification, if instead of saying take 1 pill twice a day five ways, and we said it one way and we tried to teach people what that actually means, I don't think the words themselves will necessarily be intuitively obvious, but I think it is easier to teach it if it is said one way than if it is said five ways, and that is sort of a face validity.

At this point, we do not have outcomes on that. We have got a few little studies that are underway, but not enough to be able to say there is strong outcome data on that.

DR. DAVIS: Also, we have completed four studies in the last couple of years, and the first one was published this summer, and the rest of them are in press, and that is part of the evidence that you were asking for, but there is not a lot of other evidence.

DR. PARKER: And there is not the evidence on the best practice to say this is the way to do it. On the other hand, I think it's a call for more research that allows us to begin to come up with what the best practice really is.

The stuff about language, I am not really aware that there are published studies, but I think there again, with the growing population with that issue, we do know some numbers about literacy skills among Spanish speakers.

The warn with an icon, that is an interesting one. The work that we have done is really just on how poorly understood they are, not that we know this one works. Now, some of the human factors work--and Ruth and others may be able to comment on this better than I can--but there is

a standardized use of certain icons that people do learn and do recognize, but I think right now we are in the mode of multiple including the inverted stomach I might say, that is as mirror image, that didn't even help you improve health literacy because the stomach was backwards on that one, just a little clinical point.

The things about the PPIs had me pretty concerned. You have to come up and look at these, some nice coupons in there.

DR. BRASS: Just to clarify that last point, the ones that you were showing, the ones you have, are those the ones that are FDA reviewed or not FDA reviewed?

DR. PARKER: My understanding from my reading, and this staff here at FDA could clarify this, but my understanding is that FDA does have oversight of the PPIs that are in the OTCs.

DR. BRASS: But not everything that is put in there is necessarily an FDA-approved PPI, I think.

DR. PARKER: I don't know.

DR. BRASS: I mean as far as you could put any other supplemental information they want--

DR. PARKER: It looks like they do.

DR. LEONARD-SEGAL: The PPIs are not actually explicitly discussed in the regulations. When we add an insert and it's part of approved labeling, everything that we put in that insert has to be there and has to be then copied.

If a generic product were to come along, these coupons that you are talking about, we are sitting here chatting about, because I am not sure that any of us actually knew that there was coupons being tagged on to these.

I think the regulations are totally silent about it, which is probably why they are finding their way to these inserts.

DR. BRASS: But again, just for clarification, can a sponsor put in material inside the box without your approval?

DR. LEONARD-SEGAL: No.

DR. BRASS: Dr. Fincham.

DR. FINCHAM: First of all, I really

appreciated those presentations. I can't say that I enjoyed them. They are really a bit frightening in scope and content, but just a point I think that I would like to make, that probably everybody already knows, that these products aren't just sold in pharmacies.

They are sold in food markets, they are sold in convenience stores, gas stations, the Hilton gift shop, and I think that that really needs to be stressed, that once they are made available OTC, they are made available everywhere.

Just as an aside, there is a situation that I am aware of with an insert that accompanies a prescription product for a statin drug that is an advertisement for another statin drug, and the advertisement for the other statin drug contains more information than the actual leaflet supposedly that accompanies the drug that is being prescribed.

I will send it to you. I am in your state, too, and it is a pharmacy that you are aware of. I cannot get them to do anything about it. They just will not stop it. So, it is generated by

the manufacturers, it is supported by the pharmacy involved, and we are the gullible participants when we receive this.

MS. MAYER: Thank you for a very sobering presentation. I guess I am thinking of the emotional state of people, patients, when they receive information or when they are trying to process complex and difficult health situations, and the contribution that that may play on the issues of comprehension.

I am wondering if you yourselves have done any qualitative research on that or if you can report a bit to us on what the psychological dimensions of comprehension are.

DR. PARKER: That is a good question and there are certainly researchers that have pondered that. I am not aware of any ongoing studies, only qualitative and clinical practice information, for example, I think everyone here can imagine that if instead of your body temperature right now hovering between 36.5 and 37, or whatever you happen to be, that it were hovering between 38.5 and 39.5, you

probably wouldn't have taken in the same amount of information from any of the presentations that you have heard this morning, or let's say that your resting 02 saturation, instead of being close to 100, is down in the lower 90s or mid-90s, because of your congestive heart failure or your pneumonia or whatever, but good data to sort of tell you that.

I think it is more just observational at this point to say that illness of whatever sort, aches, pains, fever, common things, the kind of things we all have, as well as more debilitating chronic illness, which we know is increasingly prevalent and associated with the consumption of prescription medication, as well as over-the-counter medication, I am sure does have an impact, but I am not aware of good studies that have really picked up or measured that, I think it would just be more again sort of the face validity of it.

MS. MAYER: One of the articles that we read gave a metric for the under-appreciation of

risk of medication versus benefit. There appears to be some psychological dimension to that. I am just trying to get ahold of how that might influence comprehension.

DR. DAY: As the author of that paper, I would like to say that I will comment further on that and perception of risk versus benefit does not just have to do with emotional, but cognitive processing and how the information that is provided, so we will come back to it.

DR. BRASS: Dr. Benowitz.

DR. BENOWITZ: I am just curious about what you would recommend while keeping things simple to provide enough information for people who want information and who can interpret it, because there is sort of two sides. Some people find a problem because they don't find the information that they want to make a judgment.

So, how do you combine those two needs?

DR. DAVIS: I think that is the real challenge. You all have made the point that many times we overwhelm people, but then some people

want further information. One of the things I have gleaned from reading all this is what do we do with the real estate, where does all that information go. It can't all go here. Perhaps all the information that someone like you would want is impossible to fit here, so then what is the value, is it diminished or add-on, as you put it, inside.

So, I think that is the question, but I do think Ruth's study shows clearly that you can really overwhelm people, and that is very true with people with low literacy. Less is more for a lot of folks

DR. DAY: But I think there is also an imperative from people like the committee who take a look at medications as they are switching from Rx to OTC, to actually, or the manufacturer, somebody has got to define what are the three things that every single person who takes this medicine must understand.

Can you tell me what they are? You could say, well, there is not evidence. Somebody has got to be able to answer that. And if you don't know

the need to know, then, you don't know what you have got to focus on.

DR. DAVIS: I think that is your greatest challenge, what is the need to know, because it, in part, depends on who is talking or what the perspective is, is it inactive ingredients, the active ingredients, the warnings, the use? Our buzz word when we are developing materials now is what is the need to know and the need to do.

DR. BRASS: I think that is an extremely important point, because what often happens is there is absolute clarity in everybody's mind as to what that is. It is just when you go around the table, it is different for everybody, and that by the time you are done with the meeting, the sponsor has an unambiguous message to get these 30 things onto the label or else. So, I think that is extremely important.

I think we will go on to our next presentation, which will be Dr. Saul Shiffman for 20 minutes.

Consumer Behavior Studies

DR. SHIFFMAN: Good morning. It is a pleasure to be here and to have the opportunity to touch on some of the issues that we all grapple with in OTC switches.

[Slide.]

I am going to be offering some definition and analysis of some of the issues that Dr. Leonard-Segal raised and some opinionated recommendations for the committee and the agency's consideration.

I am going to touch on a range of the issues that Dr. Leonard-Segal touched on in her introduction, which necessarily means that I am going to go very quickly, and I am going to focus on how we try to predict consumer behavior in the OTC environment and suggest ways that the agency, this committee, and sponsors can both focus and expand the way that they look at consumer behavior.

[Slide.]

Now, the context for all of this I think we all well understand. We have all heard that phrase that the easy OTC switches are behind us.

So, we are in an era of increasing focus for patients, for consumers on self-management and self-treatment, and we are the midst of this new OTC paradigm where OTCs are moving toward chronic and preventive uses of OTC medications, and this places greater demand on patient behavior and on the difficulty of evaluating that behavior.

[Slide.]

Let's step back and think about what the issue is when a drug comes up for an OTC switch. Usually, the pharmacology is very well understood. We are typically dealing with a drug that has been on the market and been used safely for some time, and that is safe and effective if used properly.

So, really, the question on an OTC switch is a behavioral question, how will consumers use the drug in the OTC environment and ultimately, will consumer behavior lead to safe and effective use. So, how would we ever know how consumers are going to use it.

Dr. Brass has already introduced the idea that what we try to do is simulate an OTC

environment, and there are I would say a couple of essential elements of that simulation. We create an OTC-like environment with no learned intermediary like the doctor, and instead, then, the consumer is relying on the label.

We sample consumers who we hope will represent the OTC population, people who are interested in the treatment, but who haven't been screened for medical suitability, and then finally, we essentially step back and let the consumer make the decisions about buying the drug, about using it, whether to buy it again, when to discontinue or stop, to use the simpler word.

So, these are the essential elements of the simulation, then, other aspects of the design and actual use studies, the sample size of the populations are going to differ, and I will come back to those differences.

The simulation, like any simulation, is never perfect, and one of the ways in which a simulation is different than the real world is that people are in the study and we are asking them

questions.

This raises the concern, as in any behavioral study, that the questions we ask or the way we ask them may actually influence consumer behavior by making them focus on how they are using the drug or giving them hints about how they should use it, and this is what we refer to as the problem of reactivity.

So, there is a tension always in research, and in behavioral research especially between the need to collect data and ask the right questions, on the one hand, and concern about generating reactivity.

[Slide.]

Now, our strategy in addressing this has been to limit the intensity of our data collection activities, so, for example, we try not to contact or ask questions of the participant too often, but we have to realize this has a cost.

[Slide.]

It means we get less detail than we might otherwise get, and importantly, it means that we

rely very heavily on retrospective recall, which itself is known to introduce some error and bias.

We ask very open-ended questions to avoid giving participants a sense of what the right answer is, but this, too, has a cost. It means that the questions are often ambiguous, we are not sure what the answers mean, and it therefore limits what we can learn from the testing.

This is particularly the case in the example that Dr. Leonard-Segal mentioned, which is in a self-selection study when patients make or consumers make the wrong decision, it is often very hard to understand what their thinking has been.

That is particularly true when the self-selection study is the entry face to an actual use study, because essentially, we are reluctant to probe too deeply or too sharply lest we contaminate the consumer's thinking for the actual use phase.

[Slide.]

So, a strategy that would get around this, which is used sometimes, but perhaps not often enough, is to separate the self-selection study

from the actual use study, which frees up the investigators to delve more deeply into the reason behind self-selection decisions.

Those kind of insights can both, as Dr. Leonard-Segal suggested, change the way we tally self-selection, but can also lead to improvements in labeling, because one of the issues is that we often don't understand when the label seems to fail what it is that has misfired in that process.

So, we need to find methods that find the right balance between reactivity, on the one hand, and the need to have adequate data, on the other.

[Slide.]

Now, one of the biggest challenges in switches, which again Dr. Leonard-Segal set up, is really not methodological at all, which is how to evaluate the data that comes in. Actually, use studies in particular usually generate volumes of data describing consumer behavior in relation to each and every aspect of the label.

But since the purpose is to make a risk-benefit decision, I would argue that we have

to focus on the things that matter for risk-benefit, usually, on the substantial risk. This means the sponsors and the agency have to agree explicitly and beforehand what the core issues are for the OTC switch for real risk. Then, this is what should dictate the design of the actual use study.

So, for example, if the concern is about duration of use, obviously, the study will need to last long enough to observe whether people use it beyond the label period. If the concern is about repeated cycles of use, we obviously have to run it long enough to observe that and make that the endpoint.

An implication, which I think Dr. Brass already touched on, is that no matter how much folks especially in industry may want it, there is no one-size-fits-all template for actual use studies or any study. It is going to depend on the issues and the questions for the OTC switch.

Like in any other area of research, design is driven by objectives. Now, one practice that

has already been mentioned that tends to obscure this focus on what is most important is tallying total compliance, that is, the percent of consumers who did the right thing on each and every label element, and that has two issues.

One, by counting everything equally, it tends to obscure differences in importance, but secondly, the numbers computed this way turn out to be more a function of how many warnings and directions are on the label than of consumer behavior, and let me illustrate this is a very simplistic statistical model for those of you who are statistics geeks, I am going to assume that the decisions are independent and modeled under the cumulative binomial.

[Slide.]

Basically, what you see here, this is a hypothetical where each element, each warning or direction achieves 90 percent compliance, but we have a lot of warnings and directions.

You can see, first of all, that total compliance goes down as you have more and more

directions, and, in fact, by the time you get to 25 elements, which by the way is not unrealistic, compliance is below 10 percent, total compliance, even though the consumers have actually been compliant with 90 percent for each warning.

[Slide.]

This way of tallying things also tends to obscure differences in compliance, so here is the same model, but with 80 percent compliance on any one direction or warning, and you can see that by the time you get out here, there just isn't much difference even though the difference between 80 and 90 percent can actually be quite meaningful and significant.

This way of capturing compliance tends to hide information, to obscure it, rather than to review it.

[Slide.]

Talking about compliance numbers leads to a very important question that every sponsor has asked FDA at one time or another, which is how good is good enough. I would argue that focusing on

risk-benefit suggests that there is no one-size-fits-all answers, that it will depend on the nature of the risk that is a consequence of noncompliance.

But a few comments about benchmarks. First, in looking for benchmarks, we should realize that in behavioral outcomes, one should not expect perfection. I like the title of the previous talk, that to err really is human. People are not perfect, and behavioral research methods are not perfect, and therefore, perfection is unrealistic.

But there are two benchmarks we can also look to for how to evaluate compliance, and I will take each in turn. One is compliance with warnings in other domains, and the other is behavior in the Rx domain.

[Slide.]

Let me show you. This is going to be data from an experiment in which they tested compliance with a warning and direction to wear gloves and a mask when mixing toxic chemical.

Subjects were brought into a lab and asked

to mix the chemicals. They didn't know it was a test of the warning, they thought they were being tested on how well they could mix the chemical, but, in fact, they were randomized to different warnings.

In fact, you can see here, a text-only warning versus a text warning with pictograms, and these were actually signs about a foot square, and they were tested in cluttered and uncluttered environments.

So, a cluttered environment here meant that there was lots of other stuff in the room. I think of it as kind of like my office back home. The study evaluated actual compliance with the warning. So, let me show you some of the data.

[Slide.]

This shows the effect of no warning text and pictorials in a cluttered and uncluttered environment. Let's break out what we see.

First of all, we see that warnings work. Compliance was better with a warning than without. This is like our placebo control comparison. You

see that pictorials did somewhat better than text only although in this small study, that wasn't significant.

You see that clutter matters and that does have implications as we create warnings that are more and more complex and cluttered, that they may actually impede rather than improve compliance, but the big picture is that compliance was relatively poor, that it was less than 50 percent even in the best case.

[Slide.]

Now, the investigators tested some additional interventions which enhanced the warning. One was an automated voice that articulated the warning. You can see that dramatically improved compliance presumably by capturing attention, and that suggests that we need to think hard about how to capture consumer attention in a cluttered environment like the drugstore or food store.

[Slide.]

Finally, the investigators added a strobe

light, and that, too, improved compliance, but not to perfection, to about 85 percent, which is very much in the range of compliance that we often see with OTC labels even without a strobe light. It is important not to expect perfection, it is just not realistic.

[Slide.]

Perhaps more realistic standards for looking at OTC compliance is to look at compliance with directions in Rx products. Keep in mind that when we are considering an OTC switch, the question is really how behavior will change from the Rx to the OTC environment and whether it would produce any change that might increase risk.

As Dr. Leonard-Segal said, we sometimes idealize Rx behavior. We assume that it's Marcus Welby doing everything right in a compliant patient, and we seldom have data against which to evaluate that assumption, but I want to talk to you about one example where we do have data, one that I was involved with, which was the OTC switch of nicotine patch and gum from prescription to OTC.

There, data were collected, actually required by the agency, about the use of the products in the Rx environment prior to the switch, and let me show what we learned.

[Slide.]

This shows from patient reports, and these were patients who were not enrolled in a study a priori. They got the prescription in the course of their normal care. It shows, for example, that only about two-thirds received any instructions about how to use the product, particularly striking because both of these forms, patch and gum, are rather unusual. Only 50 percent received any information about side effects.

[Slide.]

Here is an interesting one, recommending enrollment in a behavioral program was considered standard good practice for these products, but only 20 percent received a referral to behavioral treatment.

Now, importantly, while 20 percent were referred for behavioral treatment, actually, only 4

percent undertook treatment, and this is an important reminder that even in the Rx environment, it is often the patient who is making practical decisions about treatments.

In summary, this shows that less than 5 percent received all five of these elements, and, in fact, over 20 percent received none of these instructions or recommendations. So, an important benchmark as a background for considering behavior in the OTC context is to understand what the behavior is in the Rx context.

[Slide.]

Stepping back more broadly, I want to put before the committee the idea of considering more sources of data to use in predicting OTC consumer behavior. Currently, we rely almost exclusively on the simulated OTC behavior in actual use studies, but there are other sources of data that would be informative.

Certainly, one is looking at how patients use the candidate drug in the Rx environment, which would be informative about what to expect in the

OTC environment.

Looking at actual patterns of use of the drug, which sometimes drugs are OTC in other countries before they are considered here, would be informative, and that is true even when it is behind the counter as is often the case in foreign jurisdictions, that can also be informative.

Actual use patterns of similar OTCs that are already on the market again would be informative. Now, the analogies are not perfect, but neither is the analogy between simulated OTC behavior in an actual use trial and real world OTC use.

I would argue that especially when no one source of data is perfect, multiple sources of data are generally better than any one standing alone. I should say that in constructing this list, I am expressing a preference for looking at behavior rather than looking at attitudes or self-report or predicted behavior. A lot of these look at what consumers are actually doing.

Now, finally, there are some things that

can't easily be or effectively be determined prior to approval, and in those cases, post-marketing surveillance may be very important.

Now, to be clear, I am not suggesting that each of these sources of data would be required for all switches or that any one of them would be required for any particular switch, but rather, that we ought to broaden the scope of the data we look at when we are trying to predict OTC behavior, and not limit it to simulated actual use.

[Slide.]

In the same spirit, we need to expand the means that we use to reach consumers and influence, not just study, but influence their behavior.

Now, the label, of course, is the primary source of information for the OTC consumer, but as we have discussed, materials in the package can be helpful supplements to that, and things that are not at all in the package may also be important, consumer education campaigns, and so on, and even outreach through the healthcare system, and we should consider all of these avenues to create,

where needed, a coherent risk management program where there is a need based on specific concerns about OTC use.

These might be especially important to address the needs of small vulnerable populations, such as people with rare conditions that might be adversely affected by the OTC product.

Sometimes it will make sense to consider a broad array of programs as a coherent risk management program as part of the OTC switch proposition.

[Slide.]

Now, this necessarily raises the question of how such programs are to be evaluated. Now, how much evaluation is needed depends again on how important the program is to the OTC proposition especially to protecting patient safety, and if the program is considered critical to the OTC switch, then, the sponsor may need to demonstrate that with the program, the product meets the OTC standard of safety. This can be done by including the program as an element in the OTC simulation in the actual

use study, so that that evaluation incorporates the value of the program.

Now, in my view, randomized comparisons with and without the program would certainly be interesting, but really should not be necessary because really, the question we need to answer is whether with the program, the product meets the OTC standards, not necessarily to evaluate quantitatively the incremental value of the program.

Regardless, again, some programs can't be evaluated prior to approval because they can't really be modeled or simulated. Actual use studies look at consumer behavior in a kind of isolated setting, and it becomes very difficult or impossible in that artificial isolation to implement and therefore evaluate certain kinds of programs, and then we would have to do this in the real world environment.

In that case, we have to rely on them being developed through sound principles and evaluated in post-marketing surveillance, much as

is the case for risk management programs elsewhere at FDA. So, OTC is entering the era of risk management.

[Slide.]

I want to end on a note we have already struck and that you have heard before, which is that more research is needed. Now, there already is a lot of research in OTC. Whenever you have an OTC switch program, you are hearing research, but it is focused on the needs of that particular product and each risk.

What we are not doing is evaluating the broad generalizable questions that would lead to evidence-based principles, and some examples we have already come on are--someone already asked this--what is the data on wording, and so on.

[Slide.]

What we need is to develop a behavioral science, a methodological science of OTC switch, and I think that we need to create that field with the collaboration of industry sponsors, academic behavioral scientists, methodologists, and the

agency, and I think that would move the field forward.

I am hopeful that today's discussion is the beginning of that sort of focus.

Thank you for your attention. Let me stop there.

DR. BRASS: Thank you.

I would like to just probe a little bit more about this issue of the validity of the consumer-based clinical research methodologies as predictors of real world marketplace behaviors.

You recently published one example I think where the behaviors were not bad. On the other hand, there is the literature on vaginal anti-fungal products, H2 blockers, urinary tract analgesics, all of which suggests that the consumer research done pre-approval was not the least bit predictive of the aberrant behaviors that occur in the actual marketplace.

My question to you is are there strategies we can use to increase the validity of the prospective research we are doing, and second, are

we simply fooling ourselves when we do this kind of exercise.

DR. SHIFFMAN: I am aware of the studies that you mention, and I would say that they don't so much say that there isn't any validity as that there are gaps, as there always will be, that it is not perfectly predictive.

I do think that there are research strategies we can use and that the focus would be not a yea or nay, it's predictive or it isn't, but I would focus I think on which outcomes, which endpoints, in the context of which sort of actual use designs, are most predictive.

A specific strategy which requires relatively little new data collection is that we ought to be analyzing the switches that have already taken place, to look at the relationship in a more systematic way between the simulated OTC behavior and the actual behavior, and again, as you say, there have been scant examples, they do reach different conclusions, and more systematic review I think would give us a good feel for that.

DR. BRASS: Dr. Griffin.

DR. GRIFFIN: I like your suggestion of some post-marketing surveillance. I am wondering where would the onus of paying for those types of studies be.

DR. SHIFFMAN: Traditionally, they have been on the sponsor. Specifically, I have been fairly involved in the marketing of these smoking cessation products where there was quite an extensive post-marketing surveillance requirement with reporting requirement to the FDA.

DR. GRIFFIN: Secondly, is there a model where a product could have a limited market until these studies were done, so that instead of being approved for OTC use with the requirement for some post-marketing studies, where there could be that type of arrangement, but it wouldn't be released throughout the whole U.S.

DR. BRASS: Let me ask Dr. Segal to answer the question.

DR. SHIFFMAN: I will turf it to her.

DR. LEONARD-SEGAL: Thanks for the turf.

I don't think that under our current regulations, we actually have that opportunity, although I think it would be an excellent one. We have talked about in discussions over the years, whether we could do test markets, which is sort of what you are asking about, and I don't think that that venue really is there for us, at least at the present time, but I think it is a very good idea, something that it would be very nice to be able to explore.

DR. BRASS: Dr. D'Agostino.

DR. D'AGOSTINO: Thank you for the presentation, I enjoyed it very much.

I would like to go to your Slide 10. I think this illustrates tremendously a real problem. First of all, it is unlikely you will ever see data like this. This is assuming that the person has no recall and other things.

DR. SHIFFMAN: And it's all independent.

DR. D'AGOSTINO: Yes, it all independent, so most likely straight, but what it does illustrate is that if you look at a question at a

time, even if you tried to talk about what is important, and so forth, if you look at question at a time, if you have 90 percent compliance on everything, you would think that you had a wonderful study going here. Each question looked 90 percent, but underlying that there would be some real problems that people aren't putting it together.

I think that this is just as much an indictment of trying to summarize them as it is to say look at individual questions. So I think, which I think are your next couple of tables here, and charts, is that you really somehow or other have to think of the domains you are interested in and what is important.

I don't need an answer from you. I just want the audience not to think that this says somehow or other you can't summarize them. I think this is an indication that the line summary may be a problem and also individual questions, looking at individual questions, it may be even a more serious problem.

DR. SHIFFMAN: I guess I would say that the point I was trying to make is that we can't escape making decisions about priority and that summarizing the data, particularly by counting all violations, doesn't really help us escape that key decision-making.

DR. D'AGOSTINO: It may be much better than what you are showing here. It probably doesn't get down below 80 or something as opposed to going to zero.

DR. SHIFFMAN: No, agreed again. It is meant to be a simplistic, provocative model, not a realistic one.

DR. BRASS: Dr. Goldstein.

DR. GOLDSTEIN: I have an observation and a question. The observation is that restricted distribution or particularly in the OTC field have, in my experience, a long and dismal and essentially unsuccessful history, and are not likely to be, given today's travel and today's cultural exchanges, not likely to be very effective.

The question I have is--my daughter is a

teacher of English as a second language to a wide variety of immigrants--and this, by extension not only to Dr. Shiffman, but the previous speakers, what about the cultural and linguistic aspects of this?

The only language I can readily understand that is perhaps "universal," in quotes, is the pictorial language, which has been used in a variety of venues rather successfully, but could you comment on the cultural and linguistic aspects, difficulties as well as successes perhaps, in this arena?]

DR. BRASS: I am sure we will talk more about this later, but, Saul, do you want to comment?

DR. SHIFFMAN: I will make a very broad comment, which is it is an example of a broader challenge, which is that we are a diverse country, not only by language, by literacy, in every other way, and yet we only have one label, so you have to pitch it to the biggest possible group, and we have to recognize that no one label is ever going to be

perfect for everybody, but I think on the issue of linguistic and cultural sensitivity and ichnography, I have an intuition that Dr. Day is going to have something to say about that.

DR. BRASS: Dr. Benowitz.

DR. BENOWITZ: Saul, you made a strong point of concern about reactivity in these trials, and I am interested in how important a problem you think that is.

Have there been studies that have sort of broken up into subgroups and done intensive monitoring versus recall to see if they matter, and should that design be part of what a sponsor should do.

DR. SHIFFMAN: I raised the issue in part to say that it's a concern. In fact, in general, when people have done studies about reactivity, it has been much less of an issue than one would think.

I was an author on a study in which it is an extreme case where we looked at the reactivity of pain reports to assessing pain zero, three, six,

or 12 times a day, and there was no systematic trend over that, so that is one extreme.

But, in general, reactivity has been more of a hypothetical concern than an actual, and I raise it because we worry about it, but we don't study it very much, so I do think that that is something that needs more study.

I am not sure that it needs to be embedded or is best embedded in studies for switch. I think of that as a good example of the sort of background research that we need, that would inform OTC switch practice, not only it's best to do it in each particular case, but what we would really benefit from is establishing, in the general case, how bad is it, and what methods can you use to minimize it.

DR. BRASS: But certainly in other fields of behavioral research, there are examples of contamination with cueing and over-intervention that by extrapolation, makes this a not unreasonable concern.

DR. BENOWITZ: No, it's a reasonable concern, and I guess I may be implicitly looking at

the range of the reasonable. Obviously, if the way you frame your question is you took the pill twice a day, right? That's a very reactive question.

On the other hand, to give an example, when we ask is this drug appropriate for you, I am not convinced that question is so open that I am not convinced most respondents know what we are asking, and, in fact, if you look at the verbatims, the modal response is, well, because I have this headache.

It is appropriate, I have a headache, this is for headaches, and we take that to mean, oh, I read the label and I know what the indication is, I can easily imagine them looking at the interviewer quizzically and saying, "Are you stupid," yeah, it is appropriate, I have a headache, this is for headache. We get so vague that I am not sure that we are getting responses that are informative.

DR. BRASS: A quick question, Dr. Fincham?

DR. FINCHAM: Saul, just a follow up to Ralph's question about Slides 10 and 11. Are these empirically derived data you presented or just your

assumption of what might happen?

DR. SHIFFMAN: Those are purely statistical models based on a cumulative binomial, so they are not based on real data, and as Dr. D'Agostino pointed out, to simplify the model, they assumed that each decision is made independent of the others, which is unlikely in the real world.

DR. BRASS: Thank you Saul.

We will now take a break and reconvene promptly at 10:05.

[Break.]

Information Processing

DR. DAY: The topic of my presentation is about the cognitive accessibility of OTC information. I invite you to watch the screen. My comments are very integrated with the screen.

I am going to be giving you an overview of the basic approach of my research, cognitive principles, many non-OTC examples, OTC examples, some comments about generalities of the findings, and then conclusions.

The basic question is how do people

understand drug information, and the answer is with difficulty. There are many reasons for this. There is a very heavy information load, can be complex and technical information, but I am going to be focusing on the problem of cognitive accessibility.

Cognitive accessibility is the ease with which people can find, understand, remember, and use drug information, and hopefully, in a safe and effective manner.

Cognitive inaccessibility occurs whenever people have trouble doing any one or more of these things.

Research in my lab focuses on a wide range of information sources from TV ads, Internet, to hardcopy, and here are some of the examples within each category. Today, we are focusing on OTC. I am going to be commenting on hardcopy aspects, but there are definite implications for broadcast of OTC advertising, as well.

The basic approach has three phases in the research. In Phase No. 1, we do detailed cognitive

analyses. Wherever the information sources, we obtain quantitative measures and actually calculate measures of cognitive accessibility, and then we compare, for example, how accessible is the information, about the benefits versus the risks.

Then, we develop enhanced displays of exactly the same information, but based on cognitive principles to make the information more accessible, and then we do a variety of cognitive experiments to test the effects on such processes, as attention, comprehension, memory, problem solving, decision-making, behavior, and ultimately, health outcomes.

Many cognitive principles underlie this research. Many are shown on the screen there. The ones in red are the ones I will be commenting on most extensively today.

The typical one is load, how much information is too much. Most people focus on the concept of information load, the number of words, the number of pages, the number of risks, and so on, but that is not the important thing.

The really important thing is cognitive load. That is really the mental work that needs to be done to understand the information and use it. As a matter of fact, we can increase information load if we decrease cognitive load and people can do very well with it.

So, let's look at a sample experiment from another domain just to show the basic approach to this research, and it's going to be about medication schedules specifically for cases of polypharmacy.

Here is a list of medications given to an actual patient on a hospital release form. This is verbatim except it was written in handwriting, so I have cleared that up for you a little bit. This patient had difficulty knowing what to take when, and whether he had taken it, and so on. He had great difficulty.

Working with him, I realized this is a list format, and I translated it into a matrix format where we have the basic medications along the side and time zones during which a patient

might take it.

I spoke to the physician and said does this meet the spirit of your intention, could I give it to the patient, he said yes, that patient no longer had any problems taking his medication.

So we brought this real world example into the laboratory, and it used two formats, a list versus a matrix for the same schedule. There were two test conditions. In the memory condition, after people had studied the display, it was absent at the time of test.

In the comprehension condition, the display was present at the time of the test, so that means there is no memory load at all, the person just needs to look at the information and find what is needed during the testing, basic procedures people study, and then we performed many cognitive tasks.

Just looking at some of this information, when we asked people about various aspects, and I will give you more information about the questions in a moment, just to get an overview, the people

who at a random basis saw the matrix, did better in the memory tasks than the people on the list, a significant effect.

What about the comprehension condition? Now, comprehension, we would expect overall performance to go up, sure, but look, same effects with respect to format, so alternative representations.

So sometimes I will be talking about memory or comprehension today. For my purposes in looking at alternative representations, it really doesn't make a difference most times.

So, types of questions involved. Some were factual, such as what are the names of the drugs, how many total pills per day, et cetera, but some are inferential, and these are problem solving scenarios that go beyond the information given, and this is where the effect of alternative representations really makes a big difference.

Let's look at an example. If you leave home in the afternoon and will not be back until breakfast time the next day, how many inderal

should you take along? Okay. And we are going to plot the results in the same way.

Memory condition, no difference and pretty poor performance, comprehension condition. The people who have the matrix in front of them do terrificly. Okay. Big effect here. Notice the red bars. Although not quite significant, the fact that the people in the comprehension condition are doing numerically worse than the people in the memory condition, suggests that having the information in front of you can cause confusion, and not necessarily help you out, so that is something we are following up in other studies.

Now, what principles have we learned from this study? Alternative representations have consequences across cognitive tasks, such as memory and comprehension, and we need to use multiple cognitive tasks.

Now, let's look at the cognitive principle of chunking. The basic idea is you have got a bunch of stuff, some items, and some of them go together and are separate from others, and so let's

pull them apart.

So, we would put white space, if it's hardcopy, we would put silence if it's auditory.

Let's look at an example from print ads. Here is a brief summary, no chunking.

Another one, a little bit, but not purposeful.

Another one, getting better.

And that one is quite wonderful, so that you can scan, search and find things, and so on. You get the basic idea.

Let's look a little more at coding. The basic idea is once you have some chunks, let's give it a name, Name A and Name B. So, you can think of this as titles, but it not any old title, they have got to be good names, so we looked at what makes the name for something good versus bad or even misleading.

Here is a pharmacy leaflet, CMI, Consumer Medical Information, and let's look at chunking and coding within it.

Here is an excerpt from a real pharmacy

leaflet, "Tell your doctor, nurse, and pharmacist if you"--and there are various things you should talk about.

Question. You just looked at that. About how many things should you tell people, should you tell your doctor, nurse, or pharmacist? If I ask for a show of hands, most of you will probably say about 6 or 7 is the most common thing.

In fact, there is actually 15 to 20 depending on how you count, and the problem is, as shown in the box over to the side there, that there are bullets, but there are floating bullets out to the side and big blocks of text, and it is hard to see what is in that text.

So, let's look at an alternative representation where we chunk and code the information. We can now see what the categories are of things to tell and what the details are within each, so you can compare that versus this, and that is an example of chunking versus coding, making information more cognitively accessible.

Here is an example from the professional

label, the PI, and this is a real contraindications and warning section from a drug. I just replaced the drug name with Drug X to protect--well, I have just done it.

So, here we have it and let's say we need an addition about liver toxicity, what happens? Very often it just gets moved in there, so it would look like this. That makes it get lost functionally, so we could pull it out. That would be an example of chunking, but we can improve it by putting in chunking and coding to name the categories of warnings and contraindications.

Now, we can improve this even further by providing visual contrasts between the fonts for the text versus the titles. Now that we have looked at that, does that have any effect on comprehension, memory, et cetera.

So, in a study test paradigm, without going into the details, yes, indeed, chunking the information makes people get it much more readily than if it's unchunked, and the same thing for coded versus uncoded information.

So, now, let's turn to readability. Here is that same excerpt here. There is only 48 words, there is only 3 sentences, that's not much. Well, the grade level is 12 and beyond. That's where the metric stops, by the way, and there is 66 percent of passive voice sentences, which are harder to process.

Now, this morning a number of people talked about, well, what makes language different and how can we clarify. There are a lot of linguistic analyses which we can do, and here we pull things out and highlighted where all of the prepositions and connectives are, and there is more than are needed, and as a matter of fact, that leads us into the discussion of the lard factor.

Lard is extra fat in sentences which make it difficult to process the meaning, and we have laboratory studies on this. So, here is the original of a high lard factor, and now it's de-larded, and I won't go through this, but there is a formula for doing this and it makes it, pardon the expression, lots more better.

Another example of readability, and this is from the FDA Patient Information Sheets, which are growing in number, very interesting development, is of a prescription drug, and here is the section on risks.

We went through a bunch of these and looked at the grade level for text versus bullets, and there you have it. For the bullets, it is not just that they are chunked and separated, but the language is easier to comprehend because the readability grade level is lower, and, as a matter of fact, there are no passive sentences, passive voice sentences in bullets.

There is years of psycholinguistic research that shows people have problems processing, the passive voice takes longer, they make errors, and so on, and so forth. There are a lot of implications here for comprehension, memory, and ultimately, behavior.

Let's look now at the principle of location of information, and let's take a look at a very well-known cognitive principle, again over a

half century of research on this, very robust findings.

It is called the serial position effect. If you here see a bunch of items and then have to report them, typical results work like this. Percent correct is a function of location. People do better at the beginning and the end. They have trouble in the middle. That's true, it has been replicated hundreds of times.

All right. Now, let's look at location of information in the CMI, the same one I was talking about before.

People study the leaflet, they do various cognitive tasks. One question is what are the possible side effects. There is many, many more, and these are people from the real world identified from pharmacy records that they are on certain medications, and so forth.

We are going to plot percent correct of the side effects that were provided in the Side Effects section, in the top and the middle and the bottom of it in the column, and here are the

results. Classical serial position effect.

There were also some side effects put in the Precaution section, and not in the Side Effects section, and nobody got those. So, if there is going to be a side effect put in another section, it should also be in the Side Effect section itself, as well.

Going back to the classic effect, I will just mention in passing a lot of my research is on prescription drug ads on television, broadcast ads, and just about every single ad puts the risk in that region there, between about 50 and 85 percent of elapsed time, the hardest area to process the risk.

We did produce a new TV ad for hypothetical drug flu-aid, and the structure and content is like typical ads, and the purpose was to vary specific factors and observe effects on cognition, so we are trying to get some evidence for these observations, people see an ad, and now we manipulate where we locate the side effects.

We put them either in a favorable or

unfavorable location, the exact same visual auditory information, different only in the location of side effects.

Here are the results, percent correct side effects, unfavorable location very poor, much better in a more favorable location. There is 100 percent increase in what people can do simply by putting it in a better location.

Going back to readability now, within the same TV ad, and by the way, I just want to point out that readability is not the same thing as comprehensibility, however, it is very easy to measure, it does have predictive value for more complex measures, and therefore, you use it as a quick proxy for comprehensibility.

If you look at what grade level you would need to understand the benefits or side effects on a prescription drug ad, benefits about a 6th grade level, that's great, side effects about a 9th grade level, three grade levels higher, so there is an imbalance in the cognitive accessibility of the benefits versus the risks.

Just the analysis of the materials and when we do an experiment where people see an ad, or three ads, and we test them on benefits and risks, overall, they are great on benefits, such as indications, and very poor on the side effects.

Now, let's look more broadly at the issue of alternative representations. The basic idea, it's the same information provided in different displays. Now, I do know in OTC comprehension testing there are often comparison of this way versus that way, but very often the two different ways, there are other aspects that are creeping in, so it is not exactly the same information.

So, the alternative representation's principle says the exact same information, but displayed in different ways.

Let's look for side effects. We have found that when you give people a bunch of things, they can array them on a one-dimensional array, linear ordering, and they can do this for severity, high severity, low severity. We can say to people how serious would it be if you got this side effect

versus that side effect, and they can do this with quite a bit of consistency, and for frequency of occurrence and for how good would your health state be, you know, if you had any of these side effects.

So, why don't we take those findings and now make some new representations. We could, for example, take side effects and a package insert and show the mild versus life-threatening ones and put on pictograms, and so on. These would be bad pictograms to use.

Of course, nobody who has diarrhea, drowsiness, or nausea would be jumping for joy, and we certainly don't want to have the skull and crossbones, and so on. We can put in other pictograms, but there will be a wider conversation about pictograms a little bit later.

If we were to take off the words about the categories, we could give something like this in order to help out people with low literacy or limited English, so these are just some ideas about how alternative representation and the same information might be helpful.

What I am going to do is take exactly the same information, the same size, and show it in many different ways. The typical way is a paragraph way, and the typical way is possible side effects and clue. That is your standard phrase. It gives no clues about anything other than a list of things in text form.

Here, I have categorized it, and so I will show in red here that I have categorized it in three levels of severity: dangerous, worrisome, and mild. If you don't like those terms, substitute your own, like serious instead of dangerous, it doesn't matter, so three levels of severity.

Same and now in a list, and this might be a bullet list, but here now is a chunked plus coded list, and you can see how it looks. You can put it in a line now, in an arrow of increasing something or other, and you can start adding other enhancements to show the increase in seriousness.

Here now is a matrix where we have across the two underlying dimensions of side effects,

frequency of occurrence and severity, and, of course, they might not all populate in the way that I have here. There are different ways to do it.

Now, this can get a little more complex for consumers, but we have had great luck with a 2 by 2 matrix like this. As a matter of fact, this is quite relieving to patients. They say, "Gee, you know, chest pain and all these are terrible things, I am not going to take this drug. Oh, I see it is very rare." Okay, that is extremely rare, and then these other things, and so on. So, they get more of a sense of these side effects.

We have used these alternative representations, the same information for text, lists, linearity, matrix, and lots more, and I don't have time to show them all to you today, and there are cognitive consequences. Again, I mean effects on perception, attention, comprehension, memory, problem solving, search and find tasks, just finding where it is, and so on, and ultimately, behavior and health.

So, these alternative representations work

out like this. Generally, the original displays of information, and I am sort of averaging, this is a schematic in a sense, graph, that the original displays lead to relatively poor performance especially for side effects, and that is a pretty close to correct level there.

When we add various enhancements, we can double performance, as shown by the first green bar, or we can increase it even higher, up to 100 percent of performance of whatever you are looking at, for whatever measurement, simply by using well-known cognitive principles.

I think we need to think anew about when a label comprehension study doesn't work into the problems of people, they don't get it, or are the representations of the information standing in the way a little bit.

Final experiment here. Print ad and product web site, you will see things like this about the Side Effect section, and so people study the original versus the enhanced.

As you can see, consumers have a big

advantage in the enhanced, but for those of you physicians here, it is not limited to physicians. Physicians do no better than the consumers when it is presented in the original format, and they then improve with the enhanced, and this was for a drug with these physicians that they all prescribe regularly.

So, the point here is physicians and patients and everyone, they may differ in their knowledge and experience and expertise, however, they are all cognitive beings, they have the same cognitive processes, subject to the same cognitive factors.

In the OTC domain, let's look at--

DR. BRASS: I really need you to close down.

DR. DAY: Okay. I will just tell you--may I have permission for one minute to show the--

DR. BRASS: No, you are already a minute off.

DR. DAY: Okay. I apologize for running late. I have never done it before--

DR. BRASS: Thank you.

DR. DAY: --ever.

DR. BRASS: There is always a first time.

Questions for Dr. Day? Maybe I will begin by asking you to expand a little bit on the warning terminology you presented, and specifically, that I can imagine that words could have equivalent comprehension, but differential effect on motivating of heeding and behaviors.

There is a literature, for example, that suggests that danger is much more likely to motivate a response than the word warning. Could you comment on that?

DR. DAY: Yes. We use a variety of different terms, and we have compared the warning type words that you just talked about with action words, so tell your doctor immediately versus tell your doctor the next visit versus continue to monitor, and so on, and they map indirectly, so you can use those action terms instead of those scary warning terms.

DR. DAVIS: For me, I love the matrix, but

do you have any research on people with low literacy? I know people with low literacy have trouble with tables and graphs.

DR. DAY: Yes, we do test people across a wide range of literacy in different studies, and we find that all of them are helpful to a certain extent. There have been some others that we have done where we then added information load, and everybody got it. Let me put it to you this way. They get lots more than if you give them the original.

So, the low literacy people will still be lower than the high literacy people, but they are--pardon the expression--lots more better than if we gave them the original representation, and that is what I am looking at.

DR. DAVIS: But they can follow a matrix.

DR. DAY: Yes, they can follow a lot of the matrix, and if you will look on the screen, there are three alternative representations for the warnings in a Drug Facts label, there is big effects on performance simply by chunking and

coding.

DR. BRASS: Ms. Mayer.

MS. MAYER: Some of what you have presented remind me of Gil Welch's and Lisa Schwartz's work on presenting study results as a part of drug labeling, the idea of using a matrix specifically.

DR. DAY: All bullets are not the same and all matrices are not the same, so it isn't which one is the one we should use, but what are the features of them and what are the consequences likely to be for cognition.

DR. BRASS: Dr. Snodgrass.

DR. SNODGRASS: Are there some persons who are primarily extremely so auditory learners versus visual learners, and that may be a different category entirely?

DR. DAY: There is some of that, but even within, when you only do the auditory versus you only do the visual, all these principles apply across all of that, so you still get the enhancement from wherever they started from by

using these principles judiciously.

DR. BRASS: If there are no other questions, I think we will move on to Dr. D'Agostino's presentation.

DR. DAY: Thank you.

DR. BRASS: I always find it interesting, when we discuss readability and comprehension, and the briefing materials include things from the Federal Register, the single most unreadable thing in history.

Statistical Considerations

DR. D'AGOSTINO: I am going to have to jump very fast because I don't want the Chair to shut me off.

[Slide.]

The aims of the presentation, I want to identify some of the issues that have major impact on the statistical analysis and interpretation. Then, I want to discuss what I think are ways of dealing with them.

I want to give you a warning to this talk that I think all issues are ultimately statistical

issues.

[Slide.]

We are going to talk material that will cover the three types of studies that we are interested in. I am not going to focus on any one, I will go across methods that I think are applicable to all.

[Slide.]

As examples, if you keep in mind some of the OTC statin ones, and the Prilosec, where the OTC statin we actually had consultation with M.D.'s and things of that nature, and then actual physical measurements, and then the Prilosec was basically out of the patterns and understanding deal show themselves in studies.

[Slide.]

I am going to look at these different items, I will talk about a little bit about the objectives, talk about what I think are the design threats to the validity, statistical analysis issues, and then some other issues.

[Slide.]

In the study, even though it is a statistics talk, you have to go back to the objectives. There is usually many objectives, and these usually come out as questionnaire responses or behavior responses.

The first thing that I think is imperative to think about is list what the objectives are and start grouping them together and start talking about what are primary and what are secondary.

If you look at an actual use study, for example, there will be usage patterns, they will be dealing with the health professional. They will be addressing the safety, then also the efficacy. Put things into categories, chunk them, as the previous speaker just said, into categories that make sense and that you are interested in, and start understanding how you are going to measure them and how you are going to concern yourself with them, are they primary, are they secondary concerns.

[Slide.]

In dealing with the endpoints, as you start having these domains and think of this slide

as talking about dealing with domains, these groups, how are you going to analyze them, you have to start thinking about that before you run the study, are you going to look at separate questions within a domain.

If you do that, you have multiple primary outcomes, and you have an awful mess trying to interpret. The other extreme is that you somehow or other, for each domain, you set up a composite.

For example, as was mentioned earlier, the percent correct on all components, you can do some interpretation along the way, but you have to start thinking of how am I going to analyze these, how am I going to think of each question, is it going to be questions separately, or am I going to somehow or other group the questions and then analyze the grouping, and then look at the individual questions after the grouping for consistency.

Further, as you start talking about putting these studies together, you have to start saying what do I expect of these domains, what do I expect of the questions and what do I expect of the

domains. Again, this would all be done a priori.

[Slide.]

Before you get into any analysis, you have to put the study together, and there are three major threats to the validity of the study that would stop you from even going on to statistical analysis.

[Slide.]

The first one is to understand the target population versus the sample population. The target population is the population of potential users. These are the ones that you want to take the drug.

The sample population is the population that you don't actually look at in the study, and it is very important to realize that as you do each exclusion/inclusion criteria, you take yourself away from the target population and every time you go further and further away from the target population, you are limiting drastically the generalizability.

My favorite type of inclusion/exclusion

criteria is include all people who will cooperate and keep them coming to interviews. That is the typical OTC setting for individuals that you have people who will cooperate, and you have people who will keep coming back for interviews.

These things can ruin your study as you put it together, and again, as was mentioned by other previous speakers, you have to be realistic, what can you do and what can't you do, but as you put your study together, you have to ask what do I want to direct this product to and what am I actually going to use in my sample.

[Slide.]

The setting is another problem. If the setting of the study isn't a natural setting, isn't somehow or other corresponding to the actual drug use, then, how do you interpret these results?

The way you conduct the study, this thing of leading reactivity type of questions, you actually ask questions in such a way that you force people into answers, or that you lead them, or that you create such an artificial setting, these things

will ruin the validity of the study and any statistical analysis is just a whistle in the dark.

[Slide.]

I gave the one about the biases. The bottom line here is to remember that most of these studies are non-randomized, open-label studies. In most clinical trials that we tend to think of, you are going to have the randomization and the blinding to be able to reduce and eliminate biases. In these studies, you don't.

In these studies, you are aware the subjects are aware of what is going on, you are aware, the evaluators are aware, so you have to be extremely careful. You have to put all these pieces together, the target population, the setting, the other biases, these are extremely important to address.

[Slide.]

Now, given that you have put your objectives together, you have put your endpoints together, you have run a good study, we get on to analyzing the data, and we will look at some of the

questions that come up in actually analyzing the data.

[Slide.]

Again, remember in most of these studies, we don't have a control group, we have just a one-arm study, no control, no comparison group, and also remember that in most of these studies, we are dealing with measuring behavior, and quite often it is the percent that are correct responses.

[Slide.]

So, what about the data set now? Now, after you have collected all your data, you have to then ask the question what data sets am I going to actually do my analysis on.

In most of these studies there is a couple of data sets, there is possibly up to three or even more, there is the intent to treat. I am using words now analogous to the clinical trials. There is the intent to treat. These are the individuals who you entered into your study.

Some of them are going to disappear, and some of them aren't going to disappear. Some of

them are going to stay until the end of the study, and especially with the actual use studies, the intent-to-treat analysis is the one that has most validity to it.

There is a sort of analogous per protocol, these are the ones that somehow or other came to all your visits, did fill out all your forms, and what have you, and then in some of the actual use studies, I have seen people talk about the adequate users as the analysis group.

I think something analogous to the intent to treat is the valid study, the valid data set. You have to be able to understand these data sets exist, these analysis data sets exist. You have to select which ones you are going to use and how to interpret them.

[Slide.]

In these studies, the difference you get between the data sets arises out of some people dropping out and some not answering questions. You have to do everything in your power to make sure that this has been minimized. When you get the

data, you have to ask the question do I have to impute data.

I have seen when I was heavy into some of these actual use studies, that this was all missed, this was all just sort of thrown aside. It is not the case that you can't really do that, you really have to look carefully at them, and you may have to be doing some sensitivity analysis to actually get at the point of what happens as you go to different databases, what happens as you take into account individuals who dropped out, what happens as you take into account people who don't have complete data.

[Slide.]

After you have identified your analysis sets, then, you ask the question what is the procedure I am going to use. Now, most of these, if not all--although it is not most, excuse me--it is a large number use a confidence interval approach to present the data.

If you are going to use a confidence interval approach, then, I think it is imperative

that you start off by saying okay, what is the outcome variable, is it going to be the separate variable, is it going to be the composite, and what do I expect of that, what do I think of as constituting a success.

Again, this should be somehow or other stated a priori. Then, you can build a confidence interval. You get your data, compute your point estimate, and compute your confidence interval

If you started off saying, for example, 85 percent success rate is what I wanted, you get a confidence interval that goes 87 plus or minus 0.03, the 85 is in there, you have somehow or other satisfied your expectations. That is the first level and what I would call the very low level of thinking about the presentation and summarizing the data.

[Slide.]

Everybody wants to say how hard it is to compute sample sizes. It is actually trivial to compute sample sizes with confidence intervals. You are not going to dwell on the formula

obviously, but if you wanted a 3 percent margin of error, that plus or minus factor, you need about 1,000 subjects, if you wanted a 5 percent margin of error, you need 385. If formulas exist, you can get even more precise formulas.

[Slide.]

A second level of thinking of presentation is to say I want to make sure that I rule out something I think is a failure. I previously said 85 percent is what you thought of as success. Well, isn't there a number that says that if it is below this number, you must have really failed, that comprehension is really inadequate?

[Slide.]

So, if you take this approach, that you come up with a number that you want to rule out, say, it is 80 percent, you want to make sure you do better than 80 percent, you can do it as a test of hypothesis, and your null hypothesis is that your success rate is less than or equal to 80 percent versus your alternative that it is greater, and you can do a one-sided test, so you, say, use a 0.25

level of significance.

Sample size formulas and software for doing that is quite trivial, and if you have something like you want to rule out an 80 percent, and your expected value back to the beginning is 0.85, you need about 500 subjects. If you really thought you were going to do well in your study, you wanted to rule out a 0.80, 100 observations would do.

[Slide.]

Formulas exist and I am obviously not going to dwell on the formula, but formulas exist, and you can pull them out of textbooks and you can go to software to do it.

[Slide.]

Now, I think what is really a nice way of viewing this material and presenting it is to combine the two procedures, that you have a value you want to rule out, and you have a value which you expect, so you have two items that you state: What do I expect my sample to produce, and what do I want to make sure I absolutely rule out. This is

sort of borrowing from some of the non-inferiority trials' designs.

So, you get your data, you get your sample, compute your confidence internal, 95 percent confidence interval, and what you want is that the thing you want to rule out sits outside, sits on the lower end of the confidence interval, sits outside the confidence interval, and you would like your expected value to sit in, you don't want it to sit up here, so you would have some problems with that, but you would like to get your data, present a confidence interval, and in that confidence interval, you can talk about what you actually believe your data tells you, and you can actually talk about what you have ruled out of the data, and sample size methods exist to handle both of those two aspirations.

[Slide.]

Now, what I gave was sample size formulas that correspond basically to looking at one variable or one composite. What if you had multiple endpoints and you say, looking at the

variables one at a time, or you are looking at number of domains, well, in this multiplicity you have to incorporate that into your sample size, so that you take into account the fact that you are introducing the chance element, but again they can be done.

You need to worry about the sample size adjustment, you need to worry about objectives within or sample size within objectives and across sets of objectives, and I think that one talks about sample size adjustments, they don't necessarily have to be that all endpoints have the same or all domains have the same degree of importance.

You might be able to negotiate, if it's with the FDA or among your investigators, on how to weight the different domains, but you need some kind of sample size adjustment when you start looking at multiple domains, multiple endpoints.

[Slide.]

I said here that one useful way of controlling multiplicity is to think of the

composite variables. I just want to make sure that you realize what I am talking about. Within the domain you can say that I am primarily interested in saying the person did the right thing, they comprehend the label.

Well, I can look at individual questions or I can somehow or other put the questions together in a way that makes sense, and it doesn't have to be all the right answers. It can be the ultimate bottom line is the correct decision.

[Slide.]

Another issue that comes up--and this is different than your handout--another issue that comes up is how do I now present this material and how do I take into account subsets. What I have here is a possible way of presenting things. This would be the overall confidence interval, and this line here, which is not in the middle, not meant to be in the middle, that was sort of your ideal, and you could identify also what you wanted to rule out, so this is what the overall gives you.

Then, you can start asking the question

what about subgroups, and the literacy group becomes very important, literacy low versus high, and I will leave it to other experts to say how you define that, but if you got this type of result, you would be very concerned that while your overall turned out to look good, the subgroup is suffering quite a bit.

This type of analysis where you lay out subgroups and put them on the same graph with the overall, is a way of seeing consistency, and the study would be successful if your overall made its objective and these subgroups also crossed this line, that they were consistent.

They don't necessarily have to be as short as, but they somehow or other have to be consistent with, and I have deliberately drawn this not to be consistent. The age is consistent, it just makes it up here, but the literacy doesn't, and it is this way of evaluating the study to see it, and you may have to, because of the literacy issue, you may have to enrich your sample, as we have seen, or you may have to do other things to increase that

result, but this is a way of I think displaying that you didn't make it with the literacy.

[Slide.]

One other issue is a comparison of control groups. Most of these studies only have a single group. Comparison groups don't exist in them and control groups don't. If you put a comparison group, another way of putting the label together, another way of asking the questions, you would have an internal validity to the study, and it is a very useful device I think in these type of open label studies to actually make the study more acceptable and more scientifically rigorous, and you could test variations and you could look at comparison of these different study arms.

[Slide.]

Here is a list. I am not going to go deeply into any of this, but here is a list of issues that one has to think about in terms of designing and conducting and analyzing a clinical trial. With the highlights here, I have given a little bit on, but I have missed a lot of items.

We didn't talk about self-report versus verifiable data. If you want to talk about a person's cholesterol, you can ask the person, and in Framingham, we would delegate you to the Fifth Circle of Health, but doing so, you need to measure it. You have study design questions, you have length of study, when do you see an effect, how long do you have to wait, and how the trial gets monitored.

There is a lot of other things, but again I tried to focus on what I thought were the major issues that were the agenda for today.

[Slide.]

Just let me close. There are many statistical issues, none of them are really unique to consumer behavior. It is important to think out the objectives, have good samples, avoid bias, decide carefully the outcomes, and then state your expectations, and your sample sizes will work out.

Thank you.

DR. BRASS: Thank you.

I think that was a very useful

presentation to frame some of the questions we are going to be dealing with. I have always had a kind of simple perspective on this in terms of committee member decision-making.

What I always wanted to know is if you could tell me at what frequency a non-optimal behavior occurred and the consequence of that non-heeding, then, I could make some kind of quantitative assessment as to what the population public health risk of that non-heeding would be.

What you have presented are different ways to think about getting that frequency estimate. I am particularly attracted to thinking about the 95 percent confidence interval when it is applied to safety questions, that really, when it comes to safety, what you would like to be able to do is make an affirmative statement about the safety, not simply that you didn't observe any risk or that you had some point estimate that suggested there was no risk and that the top bar of the 95 percent confidence interval restricts what the risk will be.

Does that make sense? And that in terms of trying to define what is an acceptable level of a certain concomitant medication that would cause a drug interaction, it seems to me that the 95 percent confidence interval, the upper bound of that would be a much more useful parameter than the point estimate.

DR. D'AGOSTINO: Absolutely. First, if you do just the confidence interval procedure, the confidence interval procedure just sort of pegs a value, it pegs a value the way I am presenting it of your expectation, but then your margin of error could be 3 percent, 5 percent, what have you. You designed the study to be 3 percent, it turns out to be 5 percent, well, you take the consequence of that when you look at the upper value.

What I am actually suggesting wasn't so much the safety issue, or I could have given it in the context of the safety issue, but rather in the context of the efficacy, and it's the same thing. I am saying the lower bound is the important issue in that one, exactly your discussion, but here is

the lower bound.

You want the lower bound to rule out bad things.

DR. BRASS: I agree, but it seems, again, maybe it's just visceral, but in terms of the public health perspective, ensuring the safety questions is a much more compelling concern than the relative difference between an 75 or 85 percent on any of the other kinds of comprehension or direction-based issues. It is so easy to estimate the consequence of a safety concern that really nailing down its frequency has always struck me as more--

DR. D'AGOSTINO: On my Slide 7, I am saying take the issues that you are interested in, set them up in domains, and due to the lack of having large print, I have safety and efficacy seen together, but safety is a separate domain, efficacy is a separate domain. I tend not to think that efficacy is a good thing to use in these studies, but the safety is extremely important, and as a separate domain, everything I have said can be

applied directly to the safety.

The reason obviously that the focus is more on the usage pattern, and health professional contact, and so forth, that is sort of what we look at first, and I could have presented this in terms of safety and then people would have accused me of missing the point.

DR. BRASS: But I am talking about the behaviors that predict it, not--

DR. D'AGOSTINO: No, no, no, safety is the drug use.

DR. BRASS: Yes. Dr. Cantilena.

DR. CANTILENA: Ralph, can you put up Slide 26 again, please. It sort of went by quickly. Am I to understand that you are actually proposing a comparator, like an active control kind of group, in an actual use study?

DR. D'AGOSTINO: Well, if I recall correctly, one of the sponsors in the OTC statins actually had a comparison group of the Rx behavior. They compared the OTC behavior to the Rx behavior. I forget what the study was called. I am not

suggesting that you need to do that.

What I am suggesting is that by bringing in another group, and it is probably more a comparison group than a control group, you can get an internal validation of what is going on in the study, yet, hopefully, some more information about how to put a label together and how to understand the comprehension.

DR. CANTILENA: So, is that something that you would suggest that is used in the setting where you have like, you know, a new therapeutic area where you are unsure how it will actually perform in the over-the-counter?

DR. D'AGOSTINO: Yes, and if you wanted to, when we were doing the ibuprofen becoming OTC--

DR. CANTILENA: So, how would you analyze that, Ralph?

DR. D'AGOSTINO: --and aspirin was used quite often.

DR. CANTILENA: So, how you analyze that, just as it compares to the existing or as it compares to an older label, or how would you

actually statistically analyze that as like non-inferiority?

DR. D'AGOSTINO: You could do differences in proportion in terms of analyzing correct understanding of what the labels are and what the instructions are. Many of the examples that were brought up earlier today, on to OTC, they were prescription, and you see the lack of ability to understand what you are being given there. So, I think you could make those type of comparisons.

Again, I think the important message to carry away from this slide is the comparison groups bearing labels and things of that nature, not necessarily comparing it to pre-existing control group or control group active ingredient and what have you, but you could do that, and I think you would get a lot of information.

Does anybody remember offhand what was the study?

DR. BRASS: It was Bristol's, I think.

DR. D'AGOSTINO: I think they did some clever things in doing that. Whether or not it is

the way you want to design a study is another matter, but there are things you can do, and I think in particular the comparison group is a point to really emphasize.

DR. BRASS: I just want to follow up on that, because I think there are times when early use, not necessarily in the actual use study, but early use of comparison will inform the sponsor's decision-making about how to move forward.

Let me just give you one recurrent example. You and I have been at this table for 10-plus years at various times, and I still can't answer the following question. If I have a new OTC switch that interacts with coumadin, what is the best thing to put on the label to prevent people who actually use coumadin from using the OTC drug? Do I say don't use coumadin, blood thinner, warfarin, any nonprescription drug? Do I tell them they are going to bleed to death if they take this?

I still can't answer that question and I think some simple comparative trials of different ways of saying the same thing would definitively

address that for all time for all sponsors for all switches. This goes to what Dr. Shiffman was saying about generalizable data. We still don't have that kind of simple definitive information and every time a switch comes up, we have the exact same discussion about the exact same issue without any of the data to have the discussion about.

DR. D'AGOSTINO: I think also if you did something of this nature, you always have the question about what are the samples. You know you are going to drift away from the target population, there is no way out of that.

This might give us a peg in terms of making reasonable comparisons with other procedures, and we have this internal way of looking, internal validity as opposed to saying, well, what will they do outside. There is always that question about what will they do outside, but we can start ranking, we can start characterizing this procedure is better than this procedure.

DR. BRASS: That's right, but again in the example I posed, I only care how people who are on

coumadin understand that warning. If you are not on coumadin, I really don't care if you understood that warning or not.

Dr. Snodgrass.

DR. SNODGRASS: That could be generalized to besides warfarin, statins in rhabdomyolysis. You could probably pick five or 10 or 20 categories that need some sort of general approach to that.

My question also relates to the idea of comparison groups partly, and that is, is it not generally the case that in OTC type studies, that you are going to have more variability in the data set, whatever you choose to study, and how does that affect or what should you be thinking about in your design and in your statistical analyses, or is it not true that you don't have more variability, say, compared to prescription type studies?

DR. D'AGOSTINO: I think as a sort of global answer, you do have more variability, it's a broader population who is sorting things out for themselves, and these studies reflect that, and I did give--and I don't want to flash back--but I did

give the sample size formulas.

The sample size formulas are easy to obtain. The question is the one you are raising. I have decided on a particular mode of asking a question, I have decided on a particular way of following up individuals, now I have an ideal, I have a value I want to rule out, what kind of precision do I anticipate that I will get, so what kind of variability will I get, what kind of precision do I need.

The formulas don't tell you, hey, you made a mistake, you should be putting in a variance of such and such. You have to get that out of some information such as we saw in the earlier presentations, but I think you are right.

DR. SNODGRASS: It seems to me that would have a real input to your design considerations and even subpopulations. If you have got variability that is greater and then there are some groups that will respond differently than others, that may be an initial part of your design and comparison groups to try to decrease some of that variability

and also your interpretation.

DR. D'AGOSTINO: You could ask the question, for example, on this one here. The way I presented the material is that you sort of pick your overall--you have your sample, you target the population, now your sample population, and you have your set of questions, you have put them in as a composite or you have them separately, and you ask what do I expect.

Well, the sample sizes that I was giving--and this was what I was saying about the multiplicity--was sort of addressed to that, but if you started asking a multiplicity question that I need to know specifically what is going on in the low literacy, then, I have to make sure the sample size is adequate.

It is basically what I said before, 100 observations for the full, now I need 100 observations for this, and if I think there is going to be tremendous variability, I have to build that in. I don't want to be left with this, arguing with the FDA. Well, there is a lot of

consistency, look, this is what I wanted to get, it rules out the bad value, the expected value is there, and these other things, look at how nice the age did, this is just a fluke, I mean you don't want to be caught in that argument.

DR. BRASS: Dr. Griffin.

DR. GRIFFIN: I think that is where my question was going also is that it seems like this is real what you are presenting, this effect modification by literacy, that that is what all the previous speakers have talked about, and so wouldn't you just optimize your sample size by just concentrating?

If we feel like we really have to make it understandable to those people, why not just do your whole study in that population if that's the main driver of lower comprehension?

DR. D'AGOSTINO: I don't have an answer to the question. I think that I would be very uncomfortable seeing a study that only had people of low literacy in it. I would want some sort of balance.

I would probably need less sample size, but to focus on a subgroup, I think could potentially introduce a lot of other problems, that they aren't necessarily representative of the users, so I think you need a--I don't know another setting in the FDA where you could somehow or other take a very unique population and say that is enough or understanding enough for the provability.

DR. GRIFFIN: It is only a statistical question in terms of power, I guess it's a philosophic question is whether drugs--it is sort of like universal precautions--should a large portion of people--

DR. D'AGOSTINO: I know as lot of people who have very high literacy, and they think they understand what drugs are about, and are much more adventurous than probably a low literacy population as an example. If the statins went OTC, I know a lot of people who would suddenly never see their physician for statins anymore, and they aren't low literacy individuals.

DR. GRIFFIN: Exactly, so would that hurt

them to have a low literacy--I mean is it hurting anybody to gear things towards the low literacy--

DR. D'AGOSTINO: Well, I think it might be hurting them in terms of their behavior. We don't have information on how we should--potentially don't have information on how we should respond to approving it across the board.

DR. BRASS: I think this is an issue we are going to come back to, because I think the consequences of the lower performance of low literacy and the implications to that to an individual in public health issue is going to be an important theme in trying to understand how to handle that population.

Dr. Benowitz.

DR. BENOWITZ: I would like you to address, if you could a little bit, about the question of composite endpoints, because we saw from Dr. Shiffman's presentation, that if you were to ask a bunch of questions, and ask for every one to be correct, you will have a very low acceptable response rate.

If you do 25 separate analyses, you have multiple testing issues. You keep talking about the statistical approach to developing optimal composite endpoints.

DR. D'AGOSTINO: I want to make sure that you understand that Dr. Shiffman's slide was wrong in terms of the way life would work out. You would not get 25 questions that were completely independent, so move his curve up a bit in your question.

DR. BENOWITZ: But that is not important. The question, obviously, you could analyze a bunch of separate questions, or if they were related, you could develop composites, and I think this is an important question. I wonder if you could give us some statistical guidance as to how you develop optimal composite endpoints.

DR. D'AGOSTINO: I think what you are looking at is something that is sort of on the sheet here, that you would ask the question I have a number of questions to get at a particular behavior, and some of those questions are

redundant, some of them are tapping on different domains.

Now, there are a lot of statistical techniques, what they call principal components, factor analysis techniques, there are a lot of ways statistically to generate composites, so those could be applied.

What that would do is it would take a set of questions, you would get some data, you would do an analysis saying that these questions belong together, and there are lots of experts in terms of putting items together for questionnaires who could supply help on these type of things.

What I was saying was something much more gross than that, is that you have a set of questions, you have a set of questions and you know that they tap into different dimensions, basically, the sort of face validity, these questions dealing with this, this is dealing with that, and then you ask how do I describe.

I can look at each individual question and analyze each individual question and have a

tremendous multiple testing problem, or I can start saying that I expect good behavior, I expect correct behavior on all of this set of questions. Does that mean they are all right, does that mean 4 out of 5 is right? Does that mean that I have to change a little bit?

What I mean by "right," is it the absolute right answer to it or is it the right behavior. But I can do those things and when I lump them together in a more intuitive fashion, I can then analyze that component.

Now, after I analyze that component, I have to make sure every single question within that composite is consistent with the composite, sort of this type of graph.

So, to answer your question, there are very formal ways of doing this, very mathematical formal ways of putting these questions together. What I remember years ago, when we used to talk about upset stomach, we had a plethora of questions on what the consumer means by upset stomach.

They did factor analysis, we did

simple--we didn't do it, the company did it--they did surveys where they actually asked individuals the different questions about upset stomach, and they came up with domains that exist, and you can do the same here, that you can do that very mathematically, or as I say, I think you can do it in a much looser fashion and still get a useful composite.

The important thing that I was trying to get across, and I think the previous speakers, is what is important and how do I put these behaviors together, so that I am not looking at 25 individual questions to begin with but I am looking at domains, I am feeling comfortable about the domains, and I am looking at consistency within the questions that we are tapping in that domain.

DR. BRASS: Thank you very much, Dr. D'Agostino.

Dr. Wood.

Complexities of the Rx to OTC Switch

DR. WOOD: I am going to talk about a number of issues that seem to me to have come up

currently at NDAC meetings that require some discussion.

DR. BRASS: While they are playing with the slides, the discussion about whether or not pictographs and icons have utility in improving comprehension in various populations, it has been our previous discussions that, in fact, they add as much confusion as comprehension.

Could you comment on that, please?

DR. DAY: The answer is yes and no.

DR. BRASS: Thank you.

DR. DAY: Just illustrating brevity.

You can't just show pictograms to people and say what do they mean. It is very useful, but it's not the whole story. You have to do multiple tasks that tap different levels of processing.

So, after they study something with the pictograms on them, the simplest level of processing is did you see this one before, yes or no. So, did they notice it and encode it in any way. So, that is one task.

The second task is what does it mean, but

the third task is an alternative meanings task, could it also mean this, could it mean that, and so on, and we get wild results where it looks like the accuracy on a pictogram is pretty good, but either in their own meanings or alternate meanings they give something wild.

So, for example, if there is a picture of a pregnant woman with a question mark, one of the USP pictograms, it is supposed to mean are you pregnant or planning to become, et cetera, and some people will tell us she is wondering whether to tell her boyfriend, and you might not pick that up with some of the tasks, so in everything that is really important, you have got to do multiple tasks that tap different levels of processing.

DR. BRASS: Thank you.

I think we are ready. Fire away.

DR. WOOD: So, we are going to talk about three issues that have come up repeatedly at NDAC meetings: interpretation of label comprehension studies, interpretation of actual use studies, and by extension, triple-A studies, and potential use

or misuse by "non-target" populations.

[Slide.]

The problems that we see, I think, and some of them have been covered before in Ralph's talk, are that, first of all, there is seldom a predefined endpoint, and we never know what the study goal was in these label comprehension studies.

The study is done, the data is presented, and somehow we have to make sense of it with, as I will show you in a second, no consequence from these outcomes.

[Slide.]

Frequently, the data analysis seems to be a moving target , there is no agreement on passing grades. It is fine for Ralph to say we are going to agree that you need to be 85 percent right or whatever the number is, and design confidence limits around that, but if we don't even know going in what the passing grade is supposed to be, it is impossible to decide if this is reasonable or not, and these passing grades are decided often on the

seat of their pants.

There is no agreement on critical questions, and critical questions seem to me often to be defined by the outcome of the study, when it doesn't look so good, and questions that can be combined to give better results are combined to improve the outcome of the study.

There is certainly no agreement on data analysis. There is frequent post-hoc merging of what are described as "similar" questions, and frankly, that has often appeared to me as retrospective polishing of the data when the data comes in inadequately, and data manipulation in these studies is absolutely rampant.

[Slide.]

The same problems are also true in the actual use studies. I am not going to repeat them there because I am so terrified of Eric cutting me off, but there are additional comments that you should make.

These are, for instance, do the results from the actual use studies trump the comprehension

studies or vice versa. I don't know the answer to that, and it has never, to my knowledge, been discussed, and it is clearly a real issue.

I mean after all if people appear not to comprehend, but can use the drug properly, that's okay. It's like kind of testing somebody on the Driver's Handbook, and I am not doing so well on that, but they seem to be able to drive a car without smashing it very week. That is probably a better test.

Often these studies do not appear to be additive or even complementary in that they seem to test different things, they give different results, and we kind of sit around this table often and chew over that and come to no very rational decisions.

And should they have different goals, or are these goals so self-evidently different that we should be able to define them differently, and again that should be prespecified.

[Slide.]

In analyzing these data, it seems to me that we have approached this in an extraordinary

fashion, and we have not used the usual standards of data analysis, which is what Ralph was alluding to in his talk, it seems to me, and we should use the usual standards that we apply rigorously to every other data set that we look at in studies.

We should prespecify the analysis plan, we should prespecify what the critical questions and what the outcomes are, and if we are going to merge answers or outcomes, that should be prespecified, and, of course, agreed with the FDA in advance.

It should not be done when the results come in and we decide that we didn't do as well as we thought we should do. I mean this is not I am getting to re-analyze your SAT results once you didn't pass.

The study goals also need to be defined in advance, and that is something I want to come back to again, because I think one of the major deficiencies that we have got is we have done hundreds or seen hundreds of these studies, none of them have had a comparator group, therefore, we have never learned which of two methods is the best

way to do this, which of two methods gives better results, so we have never had an iterative process that has carried us forward from one experience to another, so that each time we see a study we are comparing it to something that went before, and we have gradually improved our performance as we do this.

Study goals need to be defined. We should explore how best to convey information by doing comparisons, and if one comparison clearly works, one would expect that the next study to come in would compare their approach to the previous comparison, so we would have an iterative approach that would result in improved performance, not just this throwing stuff at the wall and hoping it comes out.

We don't want to know just how badly we did, which is what we know right now. We want to know whether we did better than some accepted standard, or whether we did well enough. So standards for passing grades need to be predefined, standards for strategy for low literacy need to be

predefined. Usually, what we see is people who don't read well do worse than people who do read well. Wow, there is a sensational finding.

We don't have any standard that says how well people of low literacy have to do, and Marie was raising that question earlier on, should we define how well people who can't read well do in these studies, and should that be a passing grade.

We need to, most importantly, have consequences for unacceptable outcomes. Not once have we not approved a drug because any of these studies have been inadequate in their outcome. So there needs to be some expectation that there will be consequences for not performing well.

[Slide.]

The impression is that we just do a study and see how it comes out and then we try to justify the outcome or try to explain away poor outcomes with no consequences, and that is unacceptable.

By lacking a comparator, we have never found the best, or if we had only one comparator, the better approach, and we just find out how this

specific approach worked. We do it each time, so we are not looking at a wave of quality improvement that we are comparing to the best approach last time and gradually improving that approach as we go along, which is a great lost opportunity here, because this is frankly the best data we are ever going to see on label comprehension or any other approach.

Is there a better approach? We never know that, because we never try, and if we did try and find that this approach worked better than the previous approach that the company did, they certainly don't tell us about that, and no iterative approach which would also result in the same thing.

[Slide.]

Now, the other issue I want to discuss is one that comes up every time that at least I have been on this committee, and that is the performance of the drug in the non-target population. The issue usually comes up with questions from the committee that are phrased something like this:

"But what if this other group took it for some other disease, and would that be good or bad, or what if this other group that is not the target population misunderstands and takes it?"

These are obviously legitimate questions, but we don't have a way to respond to that in a quantitative fashion that makes any sense, at least any sense to me.

After all, we don't usually expect that drugs have to be proved safe and effective in patients who should not be taking the drug. In fact, quite the opposite, we expect that patients who are taking the drug will be those in whom we have proved the drugs to be safe or effective.

So, is there an acceptable or unacceptable level of risk that we can define in non-indicated patients? Clearly, one can go to the extreme and say in patients who would die if they took this drug if they didn't understand the label, and they shouldn't have taken it, that to me at least would be an unacceptable situation for an over-the-counter drug.

That is usually not the situation we are in, and it is usually some intermediate point that we are trying to grapple with. The supplementary question there is how would we know or how could we know if there were such an acceptable risk within the group that we are usually studying.

So, this came up in great detail when we were discussing the statins, and in which there was a lot of concern about the risk to patients who shouldn't be taking the drug and whether they would misuse it, but there was no data one way or the other to actually inform us on that.

[Slide.]

So, in conclusion, I think the first conclusion is reiterating, I think less elegantly than Ralph, and said that we need the same rigor for data analysis that we expect in other settings. That is currently not seen, in fact, anything but.

In actual use and comprehension studies, we need expectations and before we start the study as to what we want that study to show. We have no comparisons, so we have no learning or improvement

as we move along from one study to another. Clearly, this is an incredible opportunity that we are losing. I mean what we hear from the experts here they are doing studies, but what they should really be referring to are the data that have been generated at enormous expense in the studies presented to this committee.

We need them in comparisons to other groups, should there be a control group in every study, and that was asked by one of the committee already. I think there should and there are various ways we could do that.

We could have it say to an accepted label to define how it is done. It has been an unspoken accepted fact at the committee today that Rx treatment is just fine. I don't think it's just fine at all, but certainly we could compare it to comprehension of, say, another standard label, aspirin, for example, or we could compare two alternative ways to present the data in the study.

[Slide.]

Non-target groups. We need data, not

conjecture, when we are considering this, and is the potential for harm different from lack of benefit? I think it is, and it is one we need to rigorously discuss.

How does proven benefit to the target group outweigh potential harm to a non-target group? This is sort of a libertarian question perhaps, and the question I guess is does it or should it, and I am not sure we have ever discussed that, but it certainly requires some discussion.

Actual misuse versus potential misuse need to be treated differently, it seems to me, and also it seems to me that deliberate misuse is different from misunderstanding, and yet we often co-merge in the discussion these issues, and they are not the same thing at all.

[Slide.]

So, finally, I think we need to define standards that we expect to be met, and we need to define these standards before we start the studies, and the FDA need to agree on these standards before the studies are initiated.

We need to have expectations that these standards will be met. That is certainly not the case right now, and there need to be consequences when these standards are not met. If your results come in that your comprehension study did really badly, then, there should be some expectation that that isn't going to result in an approval, so that people will strive to get studies done and data presented in a way that is comprehendible by the population.

Thank you very much.

DR. BRASS: Thank you.

Questions for Dr. Wood? Yes, Dr. Benowitz.

DR. BENOWITZ: Your comment, you said that if there is a potential fatal consequence of a person not understanding a label, that that would be unacceptable, but for NSAIDs, I think there have been numerous fatalities and GI bleeds in people taking other anticoagulants, and yet there are a number of NSAIDs on the market, so how do you reconcile your point of view with how we should

manage NSAIDs?

DR. WOOD: Well, as you know, Neal, I don't believe labels work, but that is not the issue we are addressing here. The issue we are addressing is whether people understand the label.

Now, I actually think honestly that even having a label that says really clearly and that I understand really clearly this drug may cause a GI bleed, does not prevent me having a GI bleed. The reality is we have had a random event that causes a GI bleed, and knowing that you are going to have a GI bleed is nice to know, but it doesn't help you prevent the GI bleed.

I think, however, there are some issues in label comprehension studies for OTC drugs that are really important to get across. Explaining the risks to patients are one. Now, that doesn't necessarily prevent the risk, but explaining to the patient when and how they should use the drug, when and if they should take the drug, and conversely, if they should not take the drug are obviously clearly important issues that should be subject to

a patient's understanding, and if they don't understand them, the consequences should be that they shouldn't be having the drug available to take them.

Now, the reality is putting that in as a consequence would drive people to come up with labels that were a lot more comprehensible than they are right now, and that is the goal I think that we should be trying to achieve, not some sort of quasi-punishment.

DR. BRASS: Dr. Day.

DR. DAY: In the study I didn't show this morning, it was on GI bleeds and showing, first of all, people don't get it, but with some very small changes to enhance cognitive accessibility, they improve dramatically. It can be done in a very simple way within the Drug Facts Label regulations.

DR. WOOD: Yes, and I mean let's not forget this is true for Rx labels, too. You know, as we all remember the label for Vioxx was changed to say caution should be exercised in patients with heart disease. What did that mean? How many heart

attacks did that prevent?

You know, putting in a label something needs to have an action item that people do something about what's in the label, not just informing them that it's blue sky outside. You know, the answer is don't take your umbrella.

DR. BRASS: Dr. Neill.

DR. NEILL: Dr. D'Agostino and yourself both talk about studies that can be done pre-approval, and there have been some implications in conversations earlier about post-marketing surveillance, and I am anxious to hear what ideas you have regarding what needs to happen in terms of continuous improvement related to efficacy, safety after marketing, and while you are formulating your answer, I would mention that labels do work well at providing defensibility in court for sponsors and pharmacists and physicians and sometimes for patients.

DR. WOOD: I agree, but that's not the purpose unfortunately.

DR. NEILL: I agree.

DR. BRASS: Do you want to answer his question?

DR. WOOD: Yes. I mean I think labels are useful in defining the people who should take drugs, and I think we should be able to do that.

You are absolutely right, labels have become a means of defense in legal actions against various people, and that's why they are in small print and so long, but that is not the stated purpose of the label, and certainly we ought to be designing labels that convey, particularly for OTC use, convey information to consumers in a way they can understand and, most importantly, in a way they can act upon.

That is the deficiency right now, and again let me say what we are missing right now is an incredible opportunity to learn as we go and improve each label that we see here by defining what worked the last time and what worked the next time.

Eric raised that point about anticoagulants, but I don't think it is just single

issues like that, I think it applies to everything, and if we knew that, well, we have had five of these experiences before, and none of them worked very well, but this one works well, it would be a silly man or woman that took one of these five the next time.

DR. BRASS: Do you want to try your first question again? I think it got lost in your comment.

DR. NEILL: What studies should be done post-marketing to ensure that improvement happens in label comprehension, efficacy, and safety?

DR. WOOD: Sorry, I did overlook that. I am not sure how easy it would ever be, in fact, I am sure it would be impossible to do efficacy studies in the OTC setting post-marketing, because I think it would be extremely difficult to actually deal with all the variables that are in there.

It has been extremely difficult to do efficacy studies post-marketing for Rx drugs, never mind for OTC drugs.

Now, I think it would be easier perhaps to

do post-marketing surveys that were designed to determine whether patients who should not be taking drugs were taking them in the post-marketing setting, and I think that is not unreasonable.

I would hope that would result in an iterative approach to label comparisons, and we actually discussed that in detail in the advisory committee when acetaminophen was considered.

My view at that time was we should set standards, and instead of us spending hours discussing the label with zero data, we should tell the manufacturer to get the risk of acute hepatic failure down by 50 percent in four years, and they would do it.

They would come out with different strategies, and that may not be labeling. It may include, for instance, packaging the drug in smaller amounts, putting it into blister packs. You know, it is much harder to make a dramatic gesture, you know, you don't love me, you don't love me, and push 20 tablets out of a blister pack than it is to take a swig of something dramatically

out of a bottle.

So, you know, I think we could come up with iterative approaches that take into account things other than just labeling, which we are all wedded to, because it is easy to deal with.

But I think setting goals for people would result in much better outcomes than just going in and measuring things. I mean, you know, if you know you have to get into college, you do a lot better than if you just sort of say I have to satisfy dad or mum.

DR. BRASS: It depends on the family.

Dr. Snodgrass.

DR. SNODGRASS: Just a brief example to add onto the iterative post-marketing idea, and again acetaminophen. Even to this day, I have encountered in my situation, in a Hispanic area in pediatrics, parents who do not understand the limits on dose. So, if a little is good, more must be better. Their 2- or 3-year-old has fever, and we have had more than one or two cases, as you can imagine, of that. So to this day, that is still

going to recur, and the question is how do you take that subgroup and look at that.

DR. WOOD: I think again that was discussed during the acetaminophen hearing. One of the issues that was clearly on the table was removing from the market, concentrated preparations that could be lethal to the sort of children you are talking about.

I think that is a similar approach. You know, packaging can be used to improve safety in ways that are really meaningful, and we have not explored that at all. There are some Australian data that show that the rate of acetaminophen hepatic failure fell fairly dramatically with a packaging change.

DR. BRASS: I would like to pin you down just a little bit about your comment about the relative value of the label comp versus the actual use, who validates who.

You have seen examples where the label comp was actually done after the actual use study. It is almost in my mind about to the point where

the label comp is really more of a sponsor's tool than a regulatory tool, and that the behaviors in the actual use are the Phase III equivalent thinking about this, but I would be interested in your thoughts about that.

DR. WOOD: Well, I think that's right. The problem is that they are all in an artificial setting, and so somebody who is committed to education, as all of us are, believe that comprehension should be the first step to appropriate action. You know, we think about something, we read about it, we decide what the right answer is, and then we implement it is kind of the process we would like to think we go through in every-day life and that people would go through.

So, it seems to me that if we have a more iterative approach to the label comprehension study, we would actually discover what the touch points are going through that, and then the actual use study would focus on some of these issues, and you would have developed your actual use study--I am sorry--you would be using a label in the actual

use study that had learned from the process that went on before, but again that requires an iterative approach.

DR. BRASS: I agree completely from a trial design perspective, but from a regulatory perspective, when confronted with a successful AUS, that includes all the relevant--let's take the extreme example--that includes all the relevant population, all the relevant issues, and an LCS that is not optimal from a variety of perspectives that apparently are no longer relevant, are we fair in discounting the LCS at that point?

DR. WOOD: Well, I am not sure how you know, given that we have no passing grade, I am not sure how you know with such confidence that either of these studies was okay. I mean I am impressed that you have seen actual use studies that you thought were great, because I don't think I have seen that many of them.

DR. BRASS: I think I was being hypothetical.

DR. WOOD: Right. Looking forward to that

day, but seriously, I mean we don't have passing grades defined in advance. Were we to have that, taking account of the talks this morning, for instance, we might view differently whether you need every answer to be answered correctly by everybody to get the right grade, which is obviously not what we expect in real life.

Equally, we would force people I think into thinking more carefully about the questions and grouping the questions into important and unimportant questions in advance, which we don't do right now, we kind of lump them together to try and improve things when we are in a panic late in the game, and that is the problem I think.

DR. BRASS: Dr. Benowitz.

DR. BENOWITZ: The idea on the iterative process to optimize research design, it makes a lot of sense. I am just curious to know how you think that can be done. Is that something that should be sponsored by NIH's research studies or is it something that you think that FDA can do from submission to submission, sort of refined within

the FDA, how do you think that could be accomplished?

DR. WOOD: Well, I think it could be accomplished within the current system, and we have got a system, after all, where people present these studies to us, but we have no way to judge whether this study group is better or worse than one that we have looked at before.

If we always had a control group in there, we would have some measure of that. Now, we could take two extreme positions. We could say the control group should be a group interpreting a standard label that we thought was okay. I am not so keen on that actually, but that would be one control group that would give us a metric of where we sat. You know, this label is not as understandable as aspirin, and I am not saying the aspirin label is great, but I am just picking that as one example.

That would give you some feel for where you stood in the continuum of understandability.

The other approach, though, which I would

favor in addition to that, would be to have different strategies evaluated for each label, so that we try different approaches for each label.

The question that somebody asked about the pictograms, for example, would have been answered very clearly in such a study if we had looked at a study with and without pictograms.

So, I think we could do both, I think we should do both. I think it is not reasonable to come in with a result where we have no control group, we have no way in which we can hang that on our expectation for previous labels, and we think that is okay.

We don't know if this is a group that, you know, doesn't read well, or a group that does read well, or a group that was better in reading than the aspirin label group, or whatever, and that is a big problem. We would never allow that in an efficacy study, for instance, the idea that you just showed an improvement of X in heart failure patients without some comparison would be laughable, laughable.

DR. BENOWITZ: But just to follow up on the mechanics of that, would the FDA, after each submission, say we have learned this from the last submission and therefore these are the new guidelines for the next submission?

DR. WOOD: Well, I think what you would do is I remember, I think it is key that these things are prespecified and pre-agreed with the FDA before the study starts. Clearly, if you came in with a comparator study, and the comparison group was one that the FDA didn't think was state-of-the-art, they would force you to do a state-of-the-art comparison.

So, the idea would be that you would compare something to what they thought was the best practice from their previous experience, and you would be trying to do better than that hopefully, which would improve our understanding, or at least do, as Ralph says, no worse within some 95 percent confidence interval with sufficient power, which is the point I think he was trying to make, as well, that we need to know that if you say you got 85

percent of the answers right, is that enough, is that good enough or not. We don't know. We have no way of telling that right now.

DR. BRASS: For a lot of these issues, for both practical and even scientific reasons, these are best done, in my opinion, early in development programs on very focused questions, not necessarily trying to attempt the global comparison in a five-arm actual use study with 6,000 people, but really focusing on these very narrow, very important questions, which I continue to believe it is in the sponsor's interest as much as the agency's interest to get addressed well in order to come forward and make a definitive statement about the effectiveness of their approach.

DR. WOOD: I agree with that, Eric, completely, except I would say that when they get to the definitive study, it still is important to do some sort of comparison, because otherwise I don't see how anyone can ever interpret it.

If you tell me somebody got X in a test, if you don't know where that stands in some

continuum, you have no idea whether that was--

DR. BRASS: I agree with that, but I come back to the point both you and Ralph made, that it is really critical to define what are the questions at that point.

DR. WOOD: Absolutely, yes.

DR. BRASS: And in some cases, it may very well require, in some cases, it may not be as complicated, and you are really trying to simply verify that you are excluding a certain absolute rate of a certain behavior, so it becomes very study in question specific.

DR. WOOD: I am sort of merging Neal's question and yours. It would seem to me nirvana would be that the process you are describing, of the presubmission question polishing, would go on and would also be conveyed to the FDA, so that there would be some understanding as we look at this, that, well, this kind of question doesn't work very well we found before, and therefore is not a good way to go, or this kind of label doesn't work very well.

DR. BRASS: And again, this is where the self-selection study kind of bridges that, where you can again in a fair, practical way, do a self-selection study on a very focused population with a very focused comparator kind of group.

We can continue questions for Dr. Wood, but I would also like to open it up to questions to other of the presenters earlier in the morning, if there are any.

Dr. Neill, I think you had a question for Dr. D'Agostino.

Final Questions from the Committee to the Speakers

DR. NEILL: Specifically, what kind of study could help inform the durability of a point estimate? You gave us a lot of different measures that we could use to set up a pre-approval study, none of which will allow, in a post-marketing environment, FDA or this committee to monitor, if you will, changes that might imply the need to change the availability of a medication, make it available where once it wasn't, restrict its

availability where it has been.

DR. D'AGOSTINO: I am not sure I heard the question. Are you asking should there be monitoring so that what we anticipated before is realized, and that there may be changes?

DR. NEILL: With regard to just the label issue now. I guess you could frame it this way. Should sponsors be allowed to grow into comprehensibility and the converse, should FDA monitor growing out of comprehensibility or readability, is it possible for the environment to change enough that those changes require monitoring, and how do you do it?

DR. D'AGOSTINO: I certainly think before it is approved, there should be some standard, a priori stated standards that have been met by well-run studies, so that aside, then, you can ask the question should there be some monitoring if there is, you know, the sort of examples I am used to, if there is some potential safety issue, that there will be immediately post-marketing studies that are a part of the approval process to make

sure that that is happening.

So, one possibility is if we find something out about the drug in terms of safety, interaction with other drugs, safety, in and of itself, studies have to be mounted and produced to do that. I think that there is a constant reacting to what is going on the marketplace in terms of this drug versus other drugs.

As far as should we be dealing with that on a steady basis, and is the company going to deal with that on a steady basis, I am not sure that I have an answer. I mean it would be wonderful if the company would take upon itself to see, after the approval is given, that what they found in their studies was realized.

It would be nice also for them to monitor as they go into new markets, expand, and what have you, that the expectations that we had before are still being met. I could design those studies for you. I am not so sure I know what the regulations would be in terms of must they be done.

To answer the question can they be done,

they certainly can be done. They are not that hard to put together. How you would get the sponsor to do it, and what the requirements would be for the sponsor outside of the fact of some sort of safety issue that popped up, I don't really know a procedure that would be imposed on the sponsor.

DR. BRASS: Dr. Goldstein.

DR. GOLDSTEIN: Actually, my question really is to Dr. Day. I would like to hear more, perhaps you can elaborate on the art of a simple declarative sentence which seems to have been lost.

I think your suggestion had some considerable merit, as did many of the other suggestions here, but I would point out that the practical application of many of these could lead, in theory at least, to a bottle of aspirin for a headache costing $20, and none of us at this table, and certainly not the general public, wants that.

Again, I was struck by Dr. Day's proposal as a way out of this, and I think the Chairman's suggestion about early ought to be considered at least about the earlier studies, but, Dr. Day, I

like your simple declarative sentence, a lost art.

DR. DAY: Thank you very much. Actually, rather than an art--it is partly an art--but there is science behind this. There is a field called psycholinguistics where linguistics and behavioral scientists have come together to study how people process different types of studies.

What you have called the simple act of sentence is the kernel sentence, it is simple, active, affirmative, declarative, so the boy hit the ball as opposed to wasn't the ball hit by the boy, and based on some of the work of Noam Chomsky, and others in the past, we have been able to look at how well people understand the information in sentences of these different grammatical structures and how long it takes them, and so on, and so forth, and there are decades of science and scientific results behind all of this.

DR. BRASS: I would like to come back to the challenges of the low literacy population amongst our various experts in that area and the point that Dr. Griffin raised, because again it was

noted it is quite clear their performance will never be as good as the higher literacy unless we bias the test, and that whether or not the standards for performance, when developed, must apply logically to that low literacy population, that if it is a standard for the safe and effective use of the drug, and that the labeling must communicate to the typical consumer, then, it seems like you must meet that standard in that population.

But that doesn't seem right to me intrinsically. If I walked into--which I have--a pharmacy or equivalent in Japan and all the labels are in Japanese, I tend not to buy much, and that the question is that if a consumer is confronted with that situation, and you had some examples in your case, do they seek alternative sources of information that make the label requirement moot, do they simply not buy the product, which makes the safety question moot, or do they buy and misuse the product, which puts a tremendous onus on us to meet the requirement in that population.

Any specific thoughts people have on that, I think is really going to help me think about this.

Alastair.

DR. WOOD: I was misunderstood if you thought I was saying they should have the same standard. I actually think there should be different standards, but they should be predefined. I mean clearly, low literacy patients will perform differently and we saw lots of examples of that from high literacy, but the standards for the two ought to be predefined, so we don't sort of say, oh, yes, 10 percent low literacy, is that enough, I don't know.

DR. BRASS: But again apropos of Dr. Griffin's question, if, in fact, we believe there is some standard that is required to use the drug safely and effectively, in an absolute sense, that is the standard that needs to be applied to the low literacy population.

DR. WOOD: Well, I think there would be different --I mean somebody raised the issue, I

think, that you don't want to deny the use of the drug to a population that does understand it, and I think that is appropriate, however, you are really talking about a different issue. You are talking about making sure that a low literacy group is not at risk from them.

So, I think you could have a more informed approach than just saying it's an 85 percent passing grade for one group, and a 50 percent for the other. I think you say do they understand not to take it with anticoagulants, if that is the question you think is an absolutely key one, and you would move towards that.

You should have a different standard for a higher literacy group. I think that is my point.

DR. BRASS: Again, that is why I focus on this question of what do they actually do.

DR. DAVIS: One of the questions I have in the spirit of some of this discussion is what are you defining as low literacy?

DR. BRASS: Pick any definition you want and there is a population that meets it, how do I

deal with that population in trying to decide whether or not the availability in the marketplace of this drug exposes that group, however you define it, to a risk.

DR. DAVIS: I got it, and I love your example earlier, but back to Dr. Griffin. I am wondering if the population, if the health literacy survey that just came out said that 12 percent of the population was proficient, I wonder if that is kind of what we are looking at, that maybe 88 percent of the rest of us have some degree of limited literacy, are struggling with the way things are now, and there are about 12 percent are proficient.

That is another thing. I mean who is this group we are talking about, and then I understand that we have got to make sure that there is a minimal understanding.

DR. BRASS: Well, again, I think that is true, but I actually don't even know that's true. Again, when I walk into a Japanese pharmacy, I am exposed to no risk by not being able to read any of

those labels, because I am not going to buy any of them.

In contrast, if I was desperate and I thought I could understand the label and bought something that was inappropriate, that would expose me to a great deal of risk potentially, and it's that asymmetry that I think is critical to trying to define how rigorous we need from a regulatory standpoint to be able to deal with this population.

Dr. Shiffman.

DR. SHIFFMAN: I think the example you gave where you could define a population in any extreme way, I think sharpens the question, which is that if we set whatever standard it is, pick a number, we can foresee that given an extreme enough definition of a low literate population, we won't be able to meet it.

So, in the end, as is done for a variety of drugs, the committee and the regulators make a risk-benefit decision applied to the population. Obviously, we have to try to drive comprehension,

and to your point, compliant and safe use in the at-risk groups, but if we set an absolute standard for an extreme group, inevitably, it is going to deny benefit to the broader population.

I want to touch on perhaps a more tactical issue that relates to analysis that Dr. D'Agostino may want to comment on, which is, you know, what we don't see in these studies--we often see over-recruitment of populations of special interest, like low literacy groups, people with conditions--what we don't see in my experience is any sort of weighting of those when you report the whole.

So one of the dangers is that as we drive these studies to over-recruit larger groups, say, of low literacy or at-risk groups, that we are then not getting a very good estimate of the population as a whole unless some sort of re-weighting to representativeness is done, and that sounds like a small technical issue, but it has a huge impact on how we evaluate the safety and efficacy in the population as a whole.

DR. PARKER: I just would ask you to think back to the pie graph that we saw and perhaps think about targeting those of average literacy in the studies, which is kind of another way, Marie, of getting at the point that you made rather than are we oversampling the low--you know, it is sort of a framing, I would say.

My reading is that they need to be likely to be read and understood by the ordinary individual, so that would be the person of average. That would be my understanding of it, so just another framing would be to consider are we targeting those of average literacy as we approach the sampling.

Inherent in that is making sure that the sample is actually representative of those that are average. Right now I don't think that is how we think of it. I think we are clumping, you know, we are making sure we oversample the low literacy, whereas, the sampling is really to target those that are of overage. It is just another way to frame it.

I really appreciate your putting the language issue there, because I think that is something that needs to be dealt with upfront rather than something that is put on as a secondary issue. We either deal with it, we talk about dealing with it, and we have a strategy for going after it beforehand that allows us to put it upfront in our analysis, or we don't.

But I think what we are doing now is looking at it as a casual issue, and it is not.

DR. BRASS: Yes. As I said these things before, as a physician takes care of a patient population that is 40 percent non-English speaking, I have always found at these meetings not particular germane.

Dr. Snodgrass;

DR. SNODGRASS: I have a question that is not exactly low literacy, but more at sort of the interprofessional level, and, Dr. Parker, you had a slide, I think it was around 18 or 19 into your slides, and it had about 7 bottles lined up, and under each one was, for example, the first was one

capsule twice a day, and then it went through various iterations.

If a physician were to write--this is again a little out of the OTC, but it is still comprehension--a physician were to write a prescription one capsule bid, how would the next 100 pharmacists type that label?

I don't know the answer to that. Would they all type it the same? Probably not. This may be not low literacy, but it's a communication issue, and it gets into your standardization issue and how standardized should be and in what settings.

DR. PARKER: There is actually a manuscript out under draft about that very thing with some pretty--it wouldn't surprise you, but some great results. I would be glad to talk to you about it, but that is just another issue.

I think from the consumer's standpoint, you know, we are just beginning, and it's anyone's toss-up, I think, although it would be very interesting to look for consensus and say what are

the three most important things on here to the average or the ordinary customer. I actually think we need to be able to answer that.

I think that if we have some sense of consensus of what those are, we need to make sure that there is a high probability that the three most important things are understood by a higher proportion of people.

There is ancillary information that is also important and can be offered to people, but I think it's that consensus about what we consider absolutely most important. I would put directions for use in that top three personally.

I am very interested in the actual ability of people to take something as simple as take one twice a day and understand it. If you have average literacy, what is your ability to take a very common direction of use and understand it in a way that you can use it correctly.

I will tell you from my research, I cannot find good information that tells us the answer to that. We have got some studies now that we are

doing looking at it, and it is so common, and there again, comparisons that take the real information, the real labels and allow some comparisons to come up with how you are going to word very simple instructions and then some sort of standardization on the other end is I think that pathway.

DR. BRASS: That is why I always like seeing the temperature for storage so prominently displayed.

Dr. D'Agostino.

DR. D'AGOSTINO: Two comments, one about the low literacy group. I would hate to be endorsing that somehow or other they could be at bigger risk than the general population by moving their targets. I think a lot depends on what are the consequences of moving the target, what are the consequences of them, the low literacy group, not comprehending correctly and not doing what is sort of acceptable behavior. I think it is a hard question.

The next item of that is that I quite glibly stated that a priori goals should be stated

in terms of what you would expect and what you want to rule out. Those are not easy questions to answer what should be ruled out, and I think part of the deliberations on the part of the company, and certainly our evaluation, is they need to be stated or we just don't know how to interpret the sample or the results especially given these sort of one-arm studies, but they do require lots of this type of discussion.

I guess what I am worried about in my experience in the past is we start, as a panel, start asking those questions, they haven't been asked before, what I was saying, and I think other people were saying, Alastair, and so forth, is that these have to be addressed before we put the studies together.

One other question about the analysis and the over-recruiting, I think the way I was trying to address this sort of over-recruiting is that you really need to look at these important subgroups, that after you have done an overall analysis, you have to make sure there is consistency in the

subgroups, and if you have overweighted one subgroup that has some level of precision, and you underweight other groups, you may find that none of the subgroups are really consistent, they are very broad.

So, one way of addressing the question is the subgroup type analysis. Another way is a weighted analysis. My initial reaction, the weighted analysis might become a holy mess in terms of trying to deal with that, what do you match it with population proportions or something, but I do think you have to worry about these subgroups. Again, there are ways of doing it.

Thank you.

DR. BRASS: Thank you. I realize there are more questions and we will have more time for discussion this afternoon. We are very close to noon, so rather than risking time problems, I think we will go ahead and close now and we will reconvene promptly at 1 o'clock for the open public hearing.

Thank you.

[Whereupon, at 11:59 a.m., the proceedings were recessed, to be resumed at 1:00 p.m.]

A F T E R N O O N P R O C E E D I N G S

Open Public Hearing

DR. BRASS: The next part of our program involves the open public session. Before we hear from the public representatives who have requested time, I have been asked to read the following:

Both the Food and Drug Administration and the public believe in a transparent process for information gathering and decision-making. To ensure such transparency at the Open Public Hearing session of the advisory committee meeting, FDA believes that it is important to understand the context of an individual's presentation.

For this reason, the FDA encourages you, the Open Public Hearing speaker, at the beginning of your written or oral statement, to advise the committee of any financial relationship that you may have with any company or any group that is likely to be impacted by the topic of this meeting.

For example, the financial information may include a company's or a group's payment for your travel, lodging, or other expenses in connection

with your attendance at the meeting.

Likewise, FDA encourages you, at the beginning of your statement. to advise the committee, if you do not have any such financial relationships. If you choose not to address this issue of financial relationships at the beginning of your statement, it will not preclude you from speaking.

With that, I would like to ask Dr. Sansgiry to begin his presentation.

DR. SANSGIRY: My name is Dr. Sujit S. Sansgiry and I am a faculty member, as an Associate Professor and Director of Graduate Studies at the University of Houston College of Pharmacy in the Department of Clinical Sciences and Administration at the Texas Medical Center.

I just want to make a quick few remarks and also give some information about my research.

Before I start I just want to state that I don't have any conflict of interest or any funding from any pharmaceutical company or any interest groups for that matter.

I heard about this meeting just a month ago from a colleague of mine at the FDA, and I was quick to respond saying that I am interested in this research, because I have been interested in this particular area for almost now 15 years.

What I want to do is more than giving any comments, just provide a brief bibliography of some selected publications that I have done in this area of over-the-counter medications in general, including both consumer selection, consumer behavior, consumer misuse, and understanding of the over-the-counter medication labels in general.

I want to give this information I believe to Darrell Lyons, who will then make copies to the various different members of the advisory board. The reason I wasn't prepared for this was the fact that I got this information quite late, and this is my first FDA advisory board meeting, so I wasn't aware of the format.

I already had a PowerPoint presentation which I got ready with, but then I realized that I could give my statement in writing, so I thought it

would take me at least a year to write down my PowerPoint slides into a statement, so I did not go ahead and do that.

What I am going to do is give you quick brief summaries of statements of some of the research that I have done. I am going to start with my oldest study that was done way back in 1995 for my dissertation, and the reason for that is that it may solve some of the issues with respect to certain consumers that we talk about.

What we did was we looked at two different labels involvement and vividness. Involvement, what we defined it as how consumers are motivated to find information or seek information to improve their behavior. Vividness is actually defined as a stimuli that is provided to the consumer and how vivid that information is, and based on the vividness of the stimuli, how easy or difficult is it for the consumer to process that particular information.

We used a model which was developed by me and my advisor a while ago. It is called the Label

Evaluation Process Model, and what it does it goes through various stages where the consumer goes first seeking information from all the medication labels to then forming attitude about that label and the product itself.

The consumer then does go through a process of evaluation and then based on the evaluation and the attitude developed, at least to purchase intention or the behavior down the road in the future to purchase the product or not.

What we found out was both involvement and vividness, there wasn't an interaction factor for most of the variables. What was interesting is if the consumer is not involved in their purchase process, they are not going to be able to understand information, but based on their perception about the product label in general, they would decide whether they would buy it or not.

So, irrespective of whether they understand the information or not, they would still purchase the product based on their evaluation and attitude towards the product label.

With respect to vividness, again, the issue is the same thing. Vividness does increase product knowledge. It did increase product knowledge to a certain extent, but at the same time, depending on their attitude towards the product label, that would lead to purchase intention.

The other study that I wanted to quickly summarize was a study done by me and my colleagues, and I don't want to name all of them, but a few, Gauri Shringapure and Marjori Pawaska, we did a series of studies before the FDA came out with their guidelines, the 1999 guidelines about over-the-medication labels.

We did a study in 1995, then, one in the year 2001, and then again one which was just done about six months ago. What we found was manufacturers do take a certain amount of time to comply to these guidelines, but the good news for the FDA is--and I agree with Dr. Shiffman, if I said that name right, is manufacturers were compliant with only the 98 percent.

So 98 percent of the products in our study where we evaluated about 400 different products were compliant with the FDA guidelines with respect to the form-size on the Drug Facts label, so that is kudos I believe to the FDA, but at the same time we also look at the compliance issue where we really cannot get 100 percent compliance.

A couple of other studies that I want to mention, a recent study that was done with a graduate student, as well, we call it the Pills Project, and the Pills Project is actually product information leaflets.

The way we designed the pills was both in English and in Spanish, and these were simulated for currently existing over-the-counter medication products, and we chose three different products. One was for aspirin, one for I believe ibuprofen, and one for acetaminophen.

We did this study in three different consumer groups - bilinguals, Spanish-speaking only, and then English-speaking populations. What we found out was the pills that we designed, which

were basically information in English and in Spanish, with a form size of approximately 10 to 11 points, increased the various different variables that we measured in the Label Evaluation Process Model to a level where the current FDA's labels and all FDA labels did not even reach the neutral point, which is the minimum competency, which was about 50 percent, versus the pills, we were able to take that comprehension as well as the whole process model up to approximately 90 percent.

The last two projects that I want to really talk about, which I think the FDA, as well as the advisory board, should start thinking about, one is a study which we did, was called tag study. What we looked at was tags that were used, the pharmacist or different manufacturers, which are actually affixed to the over-the-counter medication labels, which block the information which is present on the over-the-counter medication labels.

The reason this is important is we found that approximately 25 percent of the products had these tags which were actually blocking relevant

information that consumers need while making decisions with respect to over-the-counter medication. This particular information has already been published and it is out there, so it is nothing that has not been published.

The last study, which I think I am not going to take much time here, but is very important, is the next phase which has happened recently is something called the peg cards. I am sure you might not understand what is a peg card or a facsimile card, but it is actually for all these products which have gone behind the counter, the pseudoephedrine and ephedrine products.

Pharmacists have started using these cards on the stores, which consumers then make decisions based on, and what we found out was that there is actually no regulation by the FDA on these cards. The fact that we can deport at this point, on a study that we recently did, and it is actually out for publication in the Journal of Pharmacotherapy, where we are reporting that about 92 percent of the peg cards that we studied, about 187 different

cards from 8 different pharmacies, we found that they were below the FDA requirement of 6 points in terms of the form-size requirement.

So, what happens in these peg cards is the pharmacist or whoever is developing these peg cards, all they do is take a copy of the original over-the-counter medication label, paste it on this peg cards, and these cards are there in the pharmacies for the consumers to make those decisions, but in reality, the consumers can't even read that information.

So, what I am trying to say is this is an area where the Nonprescription Advisory Committee may want to really think about, because it is not focusing on the 20 percent, it is focusing on the 80 percent of consumers that are not able to even read the information which is out there to make their appropriate decision.

With that, I want to thank you for my time and I greatly appreciate the information that has been provided at this meeting.

DR. BRASS: Thank you.

Our next speaker will be Julie Aker.

MS. AKER: Thank you. Good afternoon. My name is Julie Aker and I am President and CEO of Concentrics Research in Indianapolis, Indiana. Concentrics is a contract research organization with over 20 years experience in designing and conducting Phase II through IV clinical trials, switch research, such as label comprehension, self-selection, and actual use, and also in designing and conducting claims research.

We have the opportunity to work with a broad range of pharmaceutical and device companies, and to that end we see a wide range or programs that include drugs and devices, various therapeutic areas, and applications to both acute and chronic conditions.

Today, I would like to offer a perspective from our vantage point on the context of the current switch research, the goals for switch research, and the similarities and differences between reference points that we are all familiar with, clinical trials, on one hand, and real life

conditions, on the other hand.

I would like to share four of the most common challenges that we see across the industry in terms of switch and also for these common challenges I will be proposing some potential solutions for your consideration.

In terms of the context of switch research, we know in clinical research that we are understanding how the drug reacts physiologically in the person. We do this through studies, such as Phase I through III clinical trials. The drug is researched extensively prior to approval as an Rx drug.

In consumer research, there may be a need to do some additional clinical trials as the drug or the device is considered for switch. However, in consumer research, we are understanding how the person reacts behaviorally with the drug. We are trying to learn if the consumer can use the drug in a safe way to achieve benefits that exceed the risks in an unsupervised OTC environment.

In terms of the goals of consumer

research, in label comprehension we are focusing on the label, not the consumer. We are learning if the label is clear enough and strong enough to communicate to a broad range of consumers what the product is used for, the directions for use, and the warnings.

In self-selection, the focus is consumer judgment, a decision-making process that involves the judgment on two things: the product's label and the consumer's individual medical history.

In actual use, the focus is consumer behavior. In these studies, we are trying to learn if the consumer can make a safe decision on selecting the drug or not, and whether the consumer can use the drug in a safe way, realizing benefits that exceed the risk when they take it home and use it in an unsupervised OTC environment.

Consumer studies are rigorous. Let's consider clinical trials that we are very familiar with. Here are some common study procedures. We can see that a protocol is put together, an IRB is utilized, sites are selected, the subjects are

screened usually for a common condition or disease state. Informed consent is obtained, medical history or procedures are conducted. The subjects are enrolled and take the device or drug home to use.

Data is collected in some manner, follow-up visits and procedures are done, and then the paperwork and drug are collected at the end. In some cases, post-approval studies are done.

In label comprehension studies, the drug is not given to the consumer, and we do not send it home with them. We are taking the first and important step to understand if the label is strong enough to communicate key information, however, there is a protocol. An IRB is not utilized because the drug is not given. There is screening, but it is minimal except for special populations.

It may be more of a confidentialty agreement that is signed with the individual. Minimal screening is done, medical history or procedures. The subjects are enrolled and it is usually a one-day interview. The drug is not taken

home, and there is generally not post-approval studies.

In actual use, the drug is selected and taken home for the consumer to use as they normally would in an unsupervised OTC environment. We are trying to balance the needs for study integrity with the needs for a realistic trial. In these cases, there is a protocol, there is an IRB that is utilized, sites are selected, minimal screening is done. Informed consent is obtained. Medical procedures and history are minimal, and the enrollment is conducted based on the self-selection decision and the risk of that self-selection decision.

The drug or device is taken home to use as they normally would, and some form of data collection is done, a diary or other method. The follow-up visits are minimal and at the end of the trial, the paperwork and drug are collected, but in these studies, post-approval studies are rarely used.

Now, let's talk a little bit about the

procedure that we all follow when we try to investigate a new drug or device. In real life, we learn about a new drug or device through a mass advertising campaign that involves awareness, increased awareness, and education.

If we are motivated, we seek out the product and then we look at the labeling to evaluate if it's right for us, and we make a series of decisions. Is it right for me, which is a self-selection decision. Do I want to purchase it, which is a value decision. Do I want to use it, which is a comfort and convenience decision, and will I choose to comply with the label, which is a behavioral decision. There are no study procedures, of course, involved in real life.

In actual use studies, the consumer starts at a bit of a disadvantage in that there is no awareness and educational campaign that starts out, so the consumer is trying to understand the drug and device and how it might might apply to them.

If they are motivated to seek out the product, they are doing so in the context of a

study. They do, however, have the opportunity to read and evaluate the label and to answer some of the same questions that are posed in real life, is it right for me, do I want to purchase it, do I want to use it, and will I choose to comply with it.

In this case, we do have regulatory framework in place in that we need to have a protocol, informed consent, and a means for data collection. The trick here is that there must be a balance between simulating the real life experience and gathering useful data.

In terms of four common challenges, I would like to share four of the most common that we see across all switch programs. First, lack of consensus on what to measure or how to judge success. I think we heard some of these from the earlier speakers today.

Evaluating self-selection. Balancing the need for naturalistic approaches with the need to collect useful data. Minimal use of post-approval research to answer long-term questions.

We already talked earlier about what the goals are for label comprehension, self-selection, and actual use. In consumer research, everything derives from the label, so great care must be taken to simplify and target the messages on the label. Consumers can only be expected to absorb so much information. The core directions and warnings need to be focused.

We also recommend that the agency and sponsor company agree on these core messages for the label upfront, realizing that this is not intended to be an Rx label, but a targeted and consumer-friendly label that can be effective in communicating key information.

We recommend iterative testing and many sponsor companies do already do iterative testing on their label comprehension studies. We have a lot to learn from consumers particularly on a new label and in a new category.

One label version and study will generally not give sufficient insights. Iterative testing is key in label comprehension.

Finally, we need to adjust our collective thinking on how to judge success. We tend to think in terms of 90 to 100 percent being an A. The consumer behaviors and experiences are variable. We also know that compliance with Rx drugs is certainly less than 100 percent and is 30 to 50 percent in some cases.

We need to adjust our thinking about how to judge success in a switch program based on relative benefits and risks, and these will be different for every drug.

In terms of self-selection, this is a very critical part of the evaluation of a drug for switch. We are trying to understand if the consumer can make a safe judgment about whether or not the drug is appropriate for them to use personally.

We want to recommend that the response "I would ask my doctor or pharmacist" is, in fact, an acceptable and realistic response for consumers to give to a self-selection question. This infers caution is being exercised by the consumer.

Remember the consumer has not had the benefit of a full advertising and educational campaign, so they are starting at a bit of a disadvantage in understanding what this drug or device is all about and how it might apply to them. We see this dynamic particularly in First in Class switches.

Two of the most common reasons for incorrect self-selection are comprehension, the consumer did not see or understand the labeling, or a conscious override in which the consumer did see the warning, but decided to override it. We see this in cases in which the consumer has had a previous experience with the Rx drug or in which their doctor has already given them insight that it might be all right for them to use, or when the consumer is strongly motivated to treat.

The key here is not absolute right/wrong, yes/no. We need to understand the consumer's response in light of the risk of that response. We also need to understand how many consumers got it incorrect, why did they get it incorrect, what is the risk of their getting it incorrect, and do the

benefits exceed the risks.

There is an ongoing challenge in actual use studies with balancing the need for study integrity or the controlled framework, with simulating a real life environment as closely as possible, the uncontrolled framework.

What is considered naturalistic is in the eye of the beholder, and to this end, this becomes a subjective judgment. These studies are not conducted in a clinical setting with physicians, so that element is in line with the OTC environment already.

There is a limit to what can be simulated. This is a study and we need to get useful data. We submit that the true naturalistic environment incorporated into these studies is really at home when the consumer is unsupervised.

The goal of these studies should be shifted from a subjective discussion about what is and is not naturalistic to understanding the safety, the benefits, and the risks in the OTC environment when the consumer is using the drug at

home unsupervised.

For your consideration, we would recommend that the intended advertising and educational messages be allowed into these studies as they would be in real life, to build a reference framework for the consumer.

Also, we recommend that follow-up phone calls and scheduling a final visit be permitted. We do not want to focus unnecessary effort on retrieving consumers in a study or waiting an undisclosed period of time for them to show up for a final visit.

We need useful data, and to the point that was made earlier about the fact that we need conjecture, we don't need conjecture, we need data, this is very important. We do agree, however, that these interventions be minimal.

We have not historically considered post-approval studies as part of the switch program although we are starting to see this trend change. We cannot incorporate everything into an actual use trial.

We have already discussed some of the challenges in maintaining a naturalistic environment in an actual use study. A post-approval setting more closely approximates the true OTC experience. We submit that the long-term trends and patterns of use be shifted to post-approval trials, just as they are in Rx research.

In summary, we recommend simplifying, targeting and agreeing on a reasonable approach and expectations upfront with a focus on risk.

For self-selection, we recommend that "Ask a doctor, pharmacist, or healthcare provider" is, in fact, an acceptable response in a self-selection question.

Understanding the reasons, or the whys, behind the incorrect responses and behaviors and the associated risk is key.

For actual use, we recommend that setting realistic expectations for what is naturalistic is important, and that the key naturalistic element in an actual use study is the consumer's use at home

unsupervised, and that we shift long-term endpoints to post-approval.

We are confident that with good upfront communications, realistic and consistent goals and expectations based on risk and benefit, and shifting long-term questions to post-approval, that we can streamline the switch process and make it more consistent for everyone.

Mr. Chairman, thank you for the opportunity to present these perspectives to the committee for consideration.

DR. BRASS: Thank you very much.

Our next speaker will be Dr. Bierer.

DR. BIERER: Good afternoon. My name is Doug Bierer and I am a consultant to the OTC pharmaceutical industry. I am speaking on behalf of the Consumer Healthcare Products Association, which is a 125-year-old trade association. Its members account for more than 90 percent of the U.S. retail sales of OTC drugs.

CHPA also represents contract research organizations which design and conduct label

comprehension and actual use trials for industry.

I am providing perspective today because of my extensive experience in the topic being discussed today. I was the sponsor's team leader for the Rx to OTC switch of omeprazole. I served as CHPA Vice President for Technical and Scientific Affairs, and I have worked extensively in the development of the Points-to-Consider paper which was part of your briefing package.

The OTC industry and CHPA have a long history of working on consumer behavior research to understand how consumers use OTC products. CHPA has held several regulatory and scientific workshops with top consumer experts and the FDA in which we have discussed the design and the conduct of label comprehension and actual use trials.

Last year, the Association hosted a Consumer Behavior Roundtable for industry and FDA in which the participants agreed upon a framework to discuss consumer research and share perspectives on designing actual use studies. The combination of all these learnings have helped us to develop

the Points-to-Consider paper, which was provided in your background materials.

Keeping in mind the questions the FDA posed in their briefing document, industry is proposing some very very specific recommendations which we think will help improve the evaluation of OTC consumer behavior research, and if implemented, could make an immediate difference.

Primarily, these recommendations center around one basic tenet, that deciding what is acceptable and not acceptable consumer behavior should be based on one thing, the actual risk to the consumer.

Keeping in mind that consumer behavior studies are a valuable resource in understanding how consumers may use an OTC drug in that type of a setting, I would like you to consider three concepts that need to be looked at when evaluating these types of data.

First, there is no overall number for success or failure of a consumer behavior study.

Second, all noncompliance with the OTC

label is not the same and must be judged on the level of risk to the consumer.

Third, an incorrect response by a consumer may, in fact, be an appropriate behavior.

Let's look at each of these concepts. We would all like to have a set of guidelines of what magic number one would have to hit in the study to show an acceptable consumer behavior. The reality is there is none.

Let me give you an example. In a label comprehension study, if 85 percent of the people know that taking an OTC drug daily means to take the drug no less than 24 hours apart, for a drug with a low toxicity, that may be an acceptable response rate, however, if 85 percent of people in a study correctly respond to a "Do Not Use" warning statement, the 15 percent who did not respond correctly could represent a large number of consumers who could be at significant risk.

Therefore, each label statement for each drug needs to be looked at separately. Likewise, it is not meaningful to calculate an average across

all the label statements. So, we simply can't establish one overall number of acceptance for consumer behavior studies.

We know that some consumers may not follow the label perfectly. While these behaviors are not desirable, it is important to understand that not all labeling noncompliance represents the same level of risk to the consumer.

For example, if a drug is to be taken for no more than 10 days, if a consumer decides to take the drug for an additional day or two, the risk of this noncompliance is probably low. This shouldn't be given the same weight as a consumer who would take an OTC drug with a contraindicated drug. So, evaluating consumer behavior should be based on the risk to the consumer and considered on a statement by statement basis.

Lastly, an incorrect response may be an appropriate behavior. We know that some consumers may override an OTC label statement even though they understand the labeling. This response may represent an educated benefit-risk decision that

the consumer has made about the use of the OTC product.

Simply grading the response as incorrect and without understanding why does not help to understand the process that the consumer went through in making this decision, and actually discredits their decision-making.

For example, someone could decide to take a lipid-lowering drug even though they were slightly below the age limit that was stated on the label. If we probe further, we may have learned that they did this because they have an elevated cholesterol level and a family member recently died of a heart attack.

They have assessed the importance of the benefit, the consequences of the harm, and made their own value judgment. Therefore, we need to consider the medical significance of the incorrect answer or action so we can truly understand the actual risk of the consumer behavior.

So, in the same vein, we note the agency does not always accept the response, "I would ask

my doctor first" as an acceptable answer in some self-selection studies. However, we know from extensive research that consumers do talk to their healthcare professionals about a wide variety in aspects of OTC drugs. So, responses such as "I would talk to my doctor or pharmacist" shouldn't be judged as incorrect in self-selection studies even if the label does not instruct them to say so.

So, what can be done? CHPA is proposing three specific recommendations that we think will help improve the evaluation of OTC consumer behavior research, and, if implemented, could make an immediate difference.

First, pre-define OTC label elements critical to the safe use of the product.

Second, focus on the basis for consumer decisions.

Third, consider the wide variety of real-world consumer behavior data.

Let's look at each of these recommendations.

We know a lot about the safety of the

proposed OTC switch since it has been used in a prescription setting for many years. There should be no surprises even for rare events, and it is also possible to identify specific behaviors that could place consumers at potential risk in an OTC setting.

These risks, both from the Rx experience and from the specific consumer behaviors are captured in the OTC labeling and are usually in the warning statements.

So, before the behavioral research is conducted, we should be able to pre-define those OTC label elements that are critical to the safe use of the drug in the OTC setting.

Second. We need to move away from an ideal view of how and when consumers should use OTC products and really focus on the risk of consumer behavior in the basis for their decisions whether they made a conscious decision to override a label statement or would first discuss the use of the product with a healthcare professional.

We believe that consumer self-reported

information is valid and a valuable source of real data is when a consumer believes they have been told something by a physician and acts upon that.

What we are proposing is that OTC medicines should be approved if the risk to the consumer of noncompliance to the critical label elements is low and the lack of a medical intermediary doesn't present an additional risk to the consumer.

In keeping with our goal of understanding the totality of consumer behavior, we also need to consider the wide variety or real world consumer data that is available. Some of these data include consumer habits and practices and survey data in which existing consumer practices in the OTC category are measured.

These data provide valuable insights into what current habits and practices may or may not need to change with the introduction of a new OTC drug. Typical switch programs may include in-depth interviews of thousands of consumers. In addition, there are studies that look at in-market use and

educational programs targeted at specific subpopulations.

Since the effectiveness of some of these programs can only be evaluated once the product is in the marketplace, the use of such programs shouldn't preclude approval of the switch. These additional consumer data provide important information to support the safety of the use of OTC drugs.

In the past 30 years, more than 25 medicines have been switched from prescription to OTC status including many First in Class switches. Many of these switches were supported by consumer behavior data. These medicines have a consistent record of overall safety and none have been removed from the marketplace for safety reasons.

So, in conclusion, we recommend that evaluating whether consumer behavior is acceptable should be based on the actual risk to the consumer, factoring on the basis for the consumer decisions and the totality of the real world behavior data.

Thank you for considering the

recommendations of the OTC industry.

DR. BRASS: Thank you very much.

At this point, the program says we are going to take a break, but I suggest we delay that. I realize there were questions this morning that we didn't get to for some of the morning speakers.

What I would like to suggest is we move directly to a discussion of the questions because I predict that your questions for the individual speakers will be germane in the context of these questions, and will give an opportunity to give a more focused context to some of our discussion. Is the committee comfortable with that? Good.

Committee Questions/Discussion

DR. BENOWITZ: There was one thing we talked about this morning, which is an overriding issue, and maybe we should talk about that first. That is, Alastair's suggestion about iterative research design development, and the question about whether that is a possibility for FDA or someone to begin doing this in a systematic way, so that you can use information from one drug development

process to inform the next one, and can the stuff be published, and what is the way to sort of generate a coherent literature about that. I think that overrides a lot of the questions, so maybe we could talk about that first.

DR. BRASS: I think that is fair and maybe I will try to summarize some of the previous discussions and some of my perspectives, that I think it is within the agency's role to suggest when opportunities arise where more and better information could be obtained by using that kind of comparative trial design.

I think part of the discussion and purpose for the discussion today is to yield suggestions to the agency when such suggestions might be particularly helpful to the committee in evaluating switch applications.

I personally continue to believe that it is in the industry's long-term interest to develop this type of generic information and that therefore the role of whether it be individual sponsors or the trade association to help guide or even fund

those kinds of studies, either to contact specific programs or as more general consumer-based research also seems like a viable option.

But I think the key point here is to the degree we think that such data help us, not only evaluate an absolute number, but give us confidence that we are making decisions about individual research programs that our conclusions are based on really solid data, and that the approaches used have been actually optimized versus arbitrarily assigned.

Dr. Segal, do you want to comment on the role of the agency in fostering, have I been fair?

DR. LEONARD-SEGAL: I think your comments have been wonderful and I think that we would welcome academic research in this area.

I think that the practicality is that we see these applications, application by application, and so I guess that from our slow snail's pace perspective, new ideas that could come to us, that could help to optimize the information that we get, application by application, would be of value, and

then could be subsequently, we could learn and move forward to the next, but I think if there were an academic attempt to look at the overriding issues in this consumer research and publish some of that information, it would help us all the more and would expedite things.

DR. BRASS: I think that Neal is suggesting that this is so important that even from your perspective, when you are confronted with a particular hypothetical switch application that has got a hypothetical really critical question, where that information must truly be optimized, suggesting to the sponsor that the only way to truly optimize is to truly compare, and that coming to some arbitrary number may or may not be good enough, but it is certainly not optimized unless you have got something to compare it to.

Dr. D'Agostino.

DR. D'AGOSTINO: I certain advocated the idea of comparison groups. I am not sure I understand how they can be put into an application or input into a mandate for an application on a

particular product.

I think it behooves the companies, the sponsors to think of comparative groups, and they can design I think better studies, get more information by having them, but I am missing the point if we are talking about somehow or other this becomes a mandate.

Are we talking in that framework?

DR. BRASS: Well, I think I used the word "recommendation," not mandate, but let me give you a specific example, which is similar to the example I used before. That a drug switch X interacts with drug Y, and that X/Y interaction causes risk, it is very important that for this switch to be approved, that the committee have confidence that people taking drug Y won't use this drug over the counter.

Now, we can use your presentation to suggest that 90 or 95 or 98 percent confidence intervals are good enough, but, in fact, if we could comparative label studies to see what is the best way to communicate to a person on Y not to use X, that would seem to provide a strategy for

optimizing the communication message, yield generalizable knowledge about interactions with Y, and give greater confidence to a regulator or a committee that was asked to evaluate the risk of the X-Y interaction.

DR. D'AGOSTINO: I agree 100 percent with that, and that was part of what I was trying to say by the idea of comparison groups, and as I say, I think it behooves the companies and sponsors to think about that and to direct themselves exactly like you were saying, so recommendations that they make better studies, and then you are not so reliant on the 85 percent and 95 percent with the comparisons, I think those are all points we should stress.

Again, as long as we are talking about recommendations, they do make better studies, and not mandates. One other thing. I am not sure how you put together previous studies, I mean previous submissions in terms of what information you will get from previous submissions to inform this submission that's before us. I think part of what

was said is the field will learn a lot, I am not so sure I know exactly how to put that all together unless we have sort of classes of drugs as we do in some of the Rx things.

DR. BRASS: I agree with that, but I would suggest that if Y is a commonly used drug, then, the experience from the hypothetical would, in fact, be extrapolatable. Again, I specifically did use the word recommendation, and as Dr. Wood emphasized, there needs to be consequences if we want to truly motivate behavior.

The consequence is if this becomes an expectation for the quality of data we expect, and the committee expresses its discomfort with data of lower quality, that recommendation becomes a mandate.

Dr. Benowitz.

DR. BENOWITZ: I would like to follow up on an issue that Andrea just talked about in terms of getting academics involved. In fact, the data really come to FDA, they don't come to academics, and what is really needed is either the resources

of FDA to perform academic analyses or some sort of data sharing system where academics can really get the data to write the paper, and I think without that we are kind of stuck.

The industry can certainly be encouraged to make this data available, but FDA has it, but no one has access to it outside the FDA. I wonder if that is possible or if our committee can make a recommendation that that should be done, because I feel strongly that is what is needed to move the field forward.

DR. BRASS: Well, again, as you know, the FDA is under restrictions on what data they can or can't make available without sponsor's approval, but again, I would say the analogy is to the Rx world where they are still not perfect, but the increasing momentum to make data from clinical trials available publicly has gained momentum, and I think there is some movement in the OTC industry, as well, to recognize the value to the sponsor, as well to the public health in making that kind of information available.

DR. BENOWITZ: Well, can we, as a committee, make a strong recommendation that that be pursued?

DR. BRASS: Without objection, so ordered?

AUDIENCE: May I make a comment?

DR. BRASS: I am sorry, I cannot accept a comment from either you or Dr. Wood unless there is a specific question that is germane to your presentation. There was no question to you.

DR. LEONARD-SEGAL: I guess that I could also comment that perhaps the industry would want to drive some of this research, and that might be a very good way to obtain it.

The other comment with regard to mandating, we recommend trial designs, the agency recommends, and sometimes we recommend that sponsors conduct trials in ways that they may not have conduced them before, because our approach is that as we learn more about the science, and maybe we learn from past errors, or things that may not have been worked out so well or have been so sufficient in the past, that we feel that we have

the opportunity to benefit from that learning and then to move forward with new recommendations.

DR. BRASS: George.

DR. GOLDSTEIN: Just to comment. I am sure, certain, in fact, that the industry would consider anything the agency has to say carefully. However, I would advise steering clear of mandates which generally do not result in the very best of which the parties are capable.

Again, I take the opportunity to remind everyone the realities of cost.

DR. BENOWITZ: I just would ask if manufacturers are not required to make all Phase 3 data available to the public, why shouldn't OTC drug data be made available to the public?

DR. GOLDSTEIN: I don't think I can answer that here, Neal. I think that is for the industry after due deliberation to consider. I certainly don't either accept it or reject it, for that matter. Let's consider it.

DR. BRASS: Dr. Griffin.

DR. GRIFFIN: I want to bring up

post-marketing surveillance again because I think that even if the best studies are done and people feel pretty confident, that doesn't mean there won't be a change in how the drugs are used or what consumers perceive.

So, I think if we really are talking about public health, then, you can't make a recommendation without following up what is actually happening. I think there are national surveys of drug use that if you want to know if people are using drug X with drug Y, we could figure that out if there was the will to know that.

I guess I am just sort of getting at what Alastair talked about, looking at real outcomes. I mean if people are dying from Tylenol overdoses, we should know that, and that should drive some of what we do or what we require people to do.

DR. BRASS: Dr D'Agostino.

DR. D'AGOSTINO: Two points on that. One is that I think the post-marketing studies that focus, for example, on safety are extremely important, and that we sort of pay close attention

to the potential safety problems.

There were a couple of comments, and maybe it is out of turn, but I am going to make them anyway. There were a couple of comments made in some of the presentations about saving some of the long-term studies and actual use components of some of the long-term studies to Phase 4.

I would hate to get us in a situation where we give approval and then we are waiting for data to confirm that our approval was the right thing. I mean we have a lot of issues in orphan drugs and fast approvals, and so forth, which I think are very important. I don't feel that this is such a compulsion that we have to approve things before we feel very comfortable about what the consumers are going to do and the consequences.

DR. BRASS: Any other comments before we turn to the specific question? If not, we are going to begin with Question 1.

There are no clear guidelines regarding the number of people that should be enrolled into label comprehension, self-selection, and actual use

studies. Please discuss the sample size that should be used in each type of study and describe the basis for your response.

I am just going to read the (a) part, too, so we can mix the discussion.

(a) In some applications, there is a need to be assured that certain populations at risk for serious harm are excluded from using the drug. We often ask for a self-selection study in a group of these patients to assess whether they may consider using the drug. Please describe what sample size should be considered for these types of studies.

I think I am going to ask Ralph to give an initial kind of response to that to start the discussion.

DR. D'AGOSTINO: This is back to some of the comments I was making for the three types of studies. If you can state, and you should, state what are the objectives, what are the domains of interest that lead from those objectives, consumer behavior, contact with physician, appropriate use, and the same spiel holds for each of the types of

studies, but you have the objectives, the domains that cover those objectives, the types of questions that you want to ask to get information about those objectives, how you organize the questions into primary, secondary, composite, separate variables, it is obviously a long process, and then your expectations, rule out values, expectations, and then the sample size becomes easy.

The question is how do you get to all of those issues that I raised, and then if we impose--or not impose--but then if we add the comparator, which I think is a phenomenally important idea for label comprehension studies, for example, then, I think that you then have, instead of single-arm studies, you have multi-arm studies.

Again, it is going through what are the objectives, what are the domains that you will need to get at for those objectives, how do you ask questions, how do you follow behavior that really are the big issues, but the sample size is the standard computations.

So, in some sense, this question is, say,

with this spiel, they do exist, the notion that is really important is can the sponsor come up a priori with this grouping of questions and come up a priori with their expectations, their rule out type of values, and what is the justification for these. The data can't be looked at later on and then try to group questions together. It has to all be a priori specified.

Do you want more?

DR. BRASS: Do other people have thoughts?

I would like to expand upon that just a little bit, because implicit in what you have said, and which I strongly endorse, is prospective definition of the key questions to be addressed by the study and an acceptable level of performance from which you get the power analysis, which gives you the sample size, and that is true for 1 and 1(a). It doesn't matter what the specific question is.

But a behavior that I have observed amongst the program designers is a sense of obligation to design one study to answer every

conceivable question, when, in point of fact, it may be both cost effective and scientifically of higher utility to do multiple studies with focused objectives, focused target populations, and focused sample sizes that increases the ability to answer the most relevant questions.

If you four or five really important questions in ia particular switch, trying to get all five addressed in one study may simply be impossible because of the design elements required to optimize the answering of the individual question.

Is that consistent with what you are saying or does that make sense?

DR. D'AGOSTINO: Yes, it is very consistent. As I was saying, these domains, that if you had one domain of interest that you wanted to address, then, you are not going to be trapped into the multiple testing situation. You will have a study which is presumably going to be much more efficient timewise.

So those design considerations, can you

put together a study, label comprehension or actual use, where you focus on one particular question? I think the answer is yes, and I think that if you do those, you will get much more efficient.

Now the sponsor may be left at the end putting three or four studies together and then somebody coming back and saying, well, how do you know, you know, somehow or other he's run across all the studies, so it is not necessarily a simple answer saying break it up, but I think in terms of thinking what is the question you want and do I need a study to answer all these questions simultaneously is very important.

You can get really trapped if you have multiple objectives, you can get really trapped in the multiple testing components of that.

DR. BRASS: Dr. Snodgrass.

DR. SNODGRASS: If in the last many years, we have had, say, 25 switches, Rx to OTC, isn't there enough information in some of that, whether it is contained within the industry or in the FDA, to come up with maybe half a dozen prototype kind

of examples of approaches to how this--even step by step--how this might occur, and that those might be the beginning of how to consider designing some studies.

DR. BRASS: Well, I think that is not unfair except that part of the reason for this meeting is that those previous 25 have often left us with a sense of can't we do better. There may be some best practices identifiable out of those.

DR. SNODGRASS: Right, and it wasn't that they were necessarily best practices but rather using something, if it's not the exact example, something similar, and then pointing out deficits, for example, and then here is an alternative approach.

DR. BRASS: I think the committee has been very good at identifying deficits in real time, but again, you see the trap we are in, because if you compare Ralph's answer to this question, and the slide that went up that says 200 versus 100 in special groups, like it was a magic number, you can see not just the quantitative discordance, but the

actual qualitative discordance in how you even approach the problem.

DR. SNODGRASS: This is my point, though. Buried in that would be considerations of, for example, what is risk. Well, risk, one type of actual risk is different from another kind of consideration, so we spell out what that is.

DR. BRASS: Again, that is addressed in later questions, because again, in the way Ralph has answered the question, has made implicit that setting the benchmark is a step in that process.

Ralph.

DR. D'AGOSTINO: Some of the actual use studies that I have seen, you know, there was this panic search as the study went along to put questions together to get at responses, and so forth. I think if you look back, you can say, yeah, there is--and I am sure there are a number of good studies in the midst--but quite often it is the bad features that sort of hit us right away.

I believe I am saying the same thing that Eric is, is that we have seen enough of the studies

to have a sense of what are the components that would make a study that we feel comfortable with, and I think that is the approach that we are talking about right now.

DR. SNODGRASS: So, if you have some published example, (a) here is what was done and wasn't so good, (b) here is what would be considered a much greater improvement, would that not be of some value?

DR. BRASS: Yes, and again this goes back to Neal's point that getting these data into the public domain allows that kind of discussion, that kind of assessment to be made in a critical and public way.

Dr. Day.

DR. DAY: In many types of academic research, we look at the number of observations that we need and do power analyses on that. Now, we can increase the number of observations by having more people and fewer questions or vice versa, more questions and fewer people.

That is not always possible in label

comprehension studies when you care about a given question, but if you have a group of questions on a particular type of task, then, you can indeed have fewer people with more data per person.

DR. D'AGOSTINO: You are saying in a very nice way what I was trying to say, that you can talk about the number of individuals you need, and then when you start saying I want to look at a number of different questions, you have to start worrying about that, and you have those two parameters that you have to deal with.

DR. DAY: Right, but you would decide some of that in advance so later on when we look at self-selection, we might have scenario questions, if it is a factorial design, everybody is going to get all the questions, and then you can do it.

DR. BRASS: But even in label comp, you could design, as has been done, where there are five questions that address the same theme, and you use that as a more robust composite score rather than depending on a single question on each theme.

DR. DAY: That is exactly what I am

saying, thank you.

DR. BRASS: George.

DR. GOLDSTEIN: I would like to come back to Dr. Snodgrass' proposal about the 25 or more Rx to OTC switches. This is a veritable treasure trove of information under NDA, as indeed the Rx to OTC switched drugs must remain.

There is nothing to stop the agency within the limits of the law, and certainly cooperation company by company to examining that with respect to the parameters raised by Dr. D'Agostino, yourself, and others on this panel.

I think this is something we have in hand, and would not necessitate having to create, as it were, new data.

DR. SNODGRASS: Could this include any studies that didn't make it?

DR. BRASS: Again, there is issues with respect to who controls and who owns the data, and what is proprietary and what is not.

DR. DAVIS: I am sorry, I am really concrete here. Why are we asking this question?

The previous studies haven't been powered sufficiently to answer the questions?

DR. BRASS: That is the hypothetical framework. Again, we could simply say what has been done in the past is just great, and that could be the answer to the question, but I think what you are hearing is that would not be the consensus opinion of the committee.

DR. DAVIS: Right, I got that, but then one of the problems with the studies that have gone with these 25 switches is that the sample size wasn't adequate.

DR. BRASS: I don't want to get bogged down in the past, I mean looking forward based on our experiences, if we were to request how these studies should be powered, what would we recommend. If there was a number of 300, 200, 100 for special populations that we are happy with, we could say that.

But what has been suggested is that the right answer to please discuss the sample size, the sample size answer is--ready for your minutes,

Darrell--big enough, okay, where big enough is dependent on the importance of the question, the benchmark, et cetera.

DR. D'AGOSTINO: Just to comment on the studies I have seen, in reviewing them, sometimes it wasn't clear what the question that was being addressed was, never mind the power wasn't enough, but there was a sequence of questions, set of questions, they all made sense, but what is it that they were trying to get out of the study, so that the objective of the study, how they were going to look at the questions, not just the sample size or the power, but how was the study put together was not necessarily clear, and what we are trying to do now I think is to not only get to the sample size, but how do you think about the study and put it together.

DR. BRASS: Dr. Fincham.

DR. FINCHAM: I guess my answer would be it just depends. I am not trying to sound trite, but I want to get back to what Ruth brought up very early on this morning. It seems like hours ago,

and it was, but what are the three things that consumers need to know in order to use a product appropriately, and I think it varies across the board depending upon what the over-the-counter product is, whether it's a hemorrhoidal wipe or whether it's a different class of a PPI, and we know there is a difference between cimetidine and ranitidine and other drugs in that class as far as what the drug interactions are, what the outcomes are, and it really is dependent upon the individual product.

I very much enjoyed Ralph's comment about all problems are statistical in origin. I am paraphrasing you a little bit, but I think that we are really putting the cart before the horse unless we look at what it is that we want to have transmitted to the consumer, and then when that is elucidated, perhaps then we can get into this wonderful discussion about power and sample size and multi-arm studies, but the realization is again, with the realization again with what Dr. Goldstein mentioned very early on again this

morning, I don't want to pay $20 for a bottle of aspirin. It really depends on what the particular drug is and what it is we want consumers to know about for appropriate and safe use.

DR. D'AGOSTINO: That is exactly why when Eric asked me the question, I spent 10 minutes saying what you need to put together in the study, and the last sentence was sample size formulas exist.

DR. BRASS: I think we are all in agreement in that. Again, if you go back to the paper Mike Weintraub and I wrote, the very first step is identify the key messages. That is always where you have got to start. You have got to make them explicit and they have to be the basis of your prospective hypotheses, and it will vary on a drug-to-drug basis.

DR. PARKER: I think in addition to sort of the clarity about the standards for adequate sampling and adequate upfront goals, has to be not only do we have recommendations regarding that needing to happen, but in addition to the

recommendations, we have expectations that that will happen and there are consequences when it doesn't.

I would put that right upfront. It is not just that we recommend that this happen. We recommend, we expect, and there are consequences when it doesn't. So it is sort of there is a logical sequence that goes with that, which we also discussed earlier this morning, but I think that actually needs to be explicit.

The other thing we mention in many of these questions, and maybe it comes out and I have missed it, but this notion of label comprehension, self-selection, and actual use, and I would just ask if we have feelings about an ordering of those studies.

I can remember times when one came before, one came after, and whether or not we also have recommendations about the ordering or if that is going to come up in one of the other ones.

DR. BRASS: I think it will come up later. If it doesn't please ask it at the end, because it

was something we discussed and I think it is important.

Again, the only difference in my mind between 1 and 1(a) is that in general, if one is dealing with a certain population at risk for serious harm, one would expect a higher threshold, which would imply a higher sample size to get enough confidence in that answer, but it's the same generic answer that I think Ralph has articulated and others have commented on.

Are there other comments about this question? Neal.

DR. BENOWITZ: In 1(a), one question I would have is whether we should ask that individuals who are actually at risk answer these questions as opposed to people who are not. For example, if you ask a question about a pregnant woman, is it fair to ask a non-pregnant woman a question about is you are pregnant or--

DR. BRASS: Well, even if you did, whose answer would you care about more?

DR. BENOWITZ: So, should we be asking

that the people at risk actually be the ones we are studying.

DR. BRASS: I think that point is an excellent one. I think it has been said other times this morning, but making it explicit, that I think many of us feel that if you want to know how the at-risk population is going to behave or how they interpret the information, you have to go directly to them. I think that's right.

DR. D'AGOSTINO: The 1(a) might be a point where a Phase IV type of study, that your running study is monitoring, the wrong people buying the drug, I mean this is the type of thing that might be possible in that context.

DR. BRASS: I would disagree. I think this is the prototype of your other statement, that you would not want to be approving a drug without being certain that there was no safety risk, and understanding the populations at risk very well.

DR. D'AGOSTINO: No, what I am saying is that absolutely that you have to do all of that before you approve. My statement is that once it's

approved, will it suddenly start getting broad use, that these individuals start sneaking back into the use population.

DR. BRASS: So, this is an example of where there is a need for not only any specific example, but in general, that when an AUS says that--or self-selection says the adverse population is not going to use the drug, and we approve it, it would really be nice if we had some examples that showed that that decision was, in fact, an appropriate one, because this goes back to what has been said multiple times today, that these studies are all invalidated at their core, that none of them have been shown to be truly predictive or marketplace behaviors, which is actually the only thing we really care about, and none of these studies have been validated.

George.

DR. GOLDSTEIN: Just a quick observation. A local sportscaster likes to say let's look at the videotape. You have got in the files of those 25 and more Rx to OTC switches, data that at least

in an epidemiologic sense can raise signals. Again, based on the parameters I have mentioned earlier, the law and mutual cooperation, can that be examined and the signals derived from it put into good use?

DR. BRASS: I think that is a fair point, but I would also say that, in fact, partly because of this, all these concerns, the switch process has been relatively conservative, and not allowed when there was question of very high risk to be approved, and therefore, the kinds of signals you are going to get from an adverse reporting spontaneous base or something are not necessarily the best markers of label heeding.

Part of the reason for this discussion is that as we move forward and begin to consider more complicated consumer behaviors required to support more complicated OTC switches, it is going to require a higher level of confidence in the database to give regulators and this committee confidence that they can say yes, we believe consumers can do that based on the data you have

presented, and that is really a forward looking, because if we look back, the examples are quite mixed quite frankly, and focusing on those kind of spike signals may not be the most sensitive to detect.

DR. GOLDSTEIN: That is why we have rear view mirrors, Mr. Chairman, so that we can look behind us to see where to go forward.

DR. BRASS: Well, except from this car it has been taken off, you see. Let's take the most extreme example. Let's look at vaginal anti-fungal drugs, which this committee approved after great deliberation with a very thoughtful schema that said it was not to be used by women who had not previously been diagnosed with a candidal infection, because consumers couldn't diagnose it.

But what is happening in the marketplace? Ten to 20 percent of users have a previous diagnosis. Now, is that good or bad? That is not my point. My point is that even though an actual use study done in support of the NDA said this is what consumers will do, once it entered the

marketplace, once other sources of information became available to consumers, the consumer behaviors changed radically.

In this case, it may or may not be an issue, but again I can cite other examples. Our ability to say that an actual use study is going to predict consumer marketplace behavior five years post-approval is nil.

MS. MAYER: So, using your very apt example, I think what I am finding difficult about considering these questions is really I suppose their narrowness. I appreciate the need for greater rigor and for iterative studies, that makes good sense to me, but given that consumers make their decisions based not only on the labeling information, but on the complex set of information that includes, of course, direct to consumer marketing, should we not be asking for research that looks at all of those influences that lead 90 percent of women who are treating their yeast infections without ever consulting a physician to do so.

I mean there are some countervailing influences here that if we don't look at, it seems to me, we are not really getting the whole picture, and we are not going to be able to make decisions that really serve the public.

DR. BRASS: Well, in fact, unfortunately, there is no way to do pre-approval what you are asking for, and that is why optimizing the study design to increase their predictive value to the degree we can, which again had been said most of the time can only be done if we have more post-marketing data to use to cross this bridge, but you are not going to be able to recreate in a clinical trial pre-approval direct to consumer advertising. I mean that is not going to happen.

Again, the problem you identify is an important one to recognize. The solution isn't going to be to increase the change to design of actual use studies to try to incorporate more of those factors, because it is not going to be possible.

MS. MAYER: So, is there not data which

can be gathered from experiences with this while these same drugs are actually prescription drugs?

DR. BRASS: Again, there are data that can be collected about how prescribed consumers use the product. That may be informative and many sponsors have included such information in their applications, but again the point is not that additional information wouldn't be useful, but we are never going to be in a position to address these broad, open-ended concerns that are going to be associated with the marketplace.

MS. MAYER: I guess I am thinking, not that I have read it in total, but certainly read the media coverage of the Institute of Medicine Drug Safety Committee report. It is very clear I think to everybody that the public safety is at pretty significant risk and that one of the main issues that needs to be addressed is post-marketing safety.

I realize that is not what this committee is considering today, but I just didn't feel I could sit here without saying that what we are

discussing here is, of course, important, but there are larger issues that unless they are addressed, are just going to continue to overwhelm us I think.

DR. BRASS: I don't think anybody would disagree, and than again I think everybody would agree that more information and post-marketing real world use would be of value.

Dr. Day.

DR. DAY: Just a brief comment. We are doing research on OTC advertising on TV, so we have some idea what happens after switch, and the main thing is, of course, that the introduction of the idea of taking some drug comes without any statement about the risks.

Just start watching prescription versus nonprescription advertising, and if you find any risks for the OTC products, I would like to hear about it.

DR. BRASS: Which again I will remind people that the FDA does not control direct to consumer advertising for OTC products after launch.

DR. DAY: Absolutely, but it is part of

her--.

DR. BRASS: Dr. Neill.

DR. NEILL: Having heard some of the studies from prior meetings and switches, I don't recall hearing data from some of the groups which, in the future, would make me more likely to value the data that I get from sponsor with regard to certain populations that are at risk for serious harm.

We have seen a lot of Rx data. We have discussed a lot low literacy, so I don't want to belabor that, but there are many patients who have already identified themselves as being poorly adherent or compliant with Rx treatment, who may be unable to appropriately self-select, and that as a group, is not a group that I have seen analyzed within self-selection study or actual use data.

In addition, this concept of "See your physician if", or "This is appropriate if you have been previously diagnosed," has always implied access, which doesn't exist, and one group that I would like to see broken out is among those

patients without access or intermittent access.

DR. BRASS: We are focusing on the power calculation, right?

DR. NEILL: Yes, these are subgroup analyses for which I would want to see some point estimates. These are populations which may be at risk, but absent looking we won't know.

There is also a group that come to use medication now through other than OTC or extra OTC, whether it's Internet, or mail order, or trips to some other regulatory venue, and there is an entire cadre of folk who may have been within the Rx population, who either by virtue of a new self-diagnosis, which is important for conditions that right now aren't considered OTC, but which this committee I expect is going to be asked to revisit, is this an OTC condition.

These patients may be newly diagnosed since the OTC switch, and their behavior, especially given a cadre of physicians who then grow up where H2's and PPIs, et cetera, have never been anything but OTC, is going to be new behavior,

as well. So, those specific populations I think would be of interest to look at.

DR. BRASS: Andrea, are there other points related to Question 1 that we haven't addressed, that you think are important? Ralph, while Andrea is thinking.

DR. D'AGOSTINO: The question of the subgroups came up in this question here that we may want to add that there may be particular subgroups that we want to make sure have adequate powerful statements.

DR. BRASS: I think that is absolutely implicit in the question. When formulating the key messages and key behaviors, that there will be subpopulations to which those will matter.

DR. LEONARD-SEGAL: I guess that one of the things that comes to mind, which isn't actually part of this question, but it does come up to the post-marketing things that have been raised, is whether if you are going to recommend post-marketing safety information, whether we should be parsing stuff out like that with regard

to literacy versus low literacy.

DR. BRASS: Can you save that for a later question? I think it fits into later questions.

DR. LEONARD-SEGAL: Sure. It comes along.

DR. BRASS: Yes, but I think there is two other questions that will specifically get to that point.

Moving on to Question 2. Please discuss how the data from consumer studies should be presented for interpretation with regard to point estimates, confidence intervals, or statistical measures.

(a) Can a threshold of success be defined where anything above the threshold is considered some guarantee that the sponsor met the standard for switch? Please discuss when this should be considered, for what types of studies and how we should determine at what level of success.

I think to a large degree, Ralph, you actually incorporated the answer to this in your first, but perhaps you would like to just reiterate the focus points.

DR. D'AGOSTINO: Again, what I was trying to do in the presentation is to address a question like this where you again, if you come into the study with expectations, and I was saying that you have expectations and then you also have values that are just unacceptable.

If you come into the study with those two ingredients, you can perform a rest of hypothesis, standard test of hypothesis, but you can also present it very nicely as a confidence interval, showing that the confidence interval excludes the things you don't want, it possibly includes the things you do want, and it makes a very nice visual presentation, and once people get used to it, it is a very sharp presentation, and then you can also do this with a subgroup.

I think that using things like confidence intervals with the understanding that sits behind them is these prespecified numbers and that little sort of forest plot that I gave with the subgroup, so it would make a very nice presentation for these actual use studies.

DR. BRASS: And you would obviously implicit in that answer say yes, thresholds can be defined, and as you previously said, the specific threshold would depend on the specific question.

DR. D'AGOSTINO: Yes, and I think it is essential. If you just go in the study and say I am going to just see what the data produces, then, you have got an exploratory study, you don't have a study for real confirmation.

DR. BRASS: Let me ask you what I think might be an unfair question. Given how studies have previously been powered, and where the degree they have been powered is to get a certain power to get the point estimate above a certain value, if one wanted to get the 95 percent confidence interval to that value, in general, the studies would need to be larger. Is that true?

DR. D'AGOSTINO: Yes, and again in terms of the way I was presenting it, it is sort of like what is your null hypothesis, and then what is your expected alternative, so both of those come in. The null hypothesis is the value you don't want to

see, you want to do better than, and the sort of alternative is what you are really expecting.

I think in terms of the confidence interval, when you come to what you are expecting, you want the expected value to be in the confidence interval, and you want a nice, tight confidence interval, so that you are saying it with some real comfort and assurance.

Let me just say one other thing here also. If you go to the comparative statements, trials where you have comparison groups, everything that I had for one sample holds for two samples as machinery for presentation and testing also.

DR. BRASS: But in point of fact, one could be a lot more exploratory to take advantage of the comparator trial in trying to optimize without having to say that. I don't need to beat my 0.05 to say that something is better than something else for practical purposes. Is that--

DR. D'AGOSTINO: It is possible, but you have to be careful. If you are going to send it to this panel, you would probably want to make sure or

have a good sense that you know which of the arms you think is going to be the winner and pilot data to endorse that.

DR. BRASS: Well, particularly if it's late, but I think if it's early, it could be much more exploratory.

DR. D'AGOSTINO: Then, it is exploratory. We haven't really separated, we have sort of talking about the study that comes here, but you can do some pilot studies, nice small studies to see where you should be going with the sort of final study that is presented for approval.

DR. BRASS: Let me just reiterate my earlier comment that I think the use of the 95 percent confidence interval is extremely useful when thinking about safety issues. When you are trying to make an affirmative statement that excludes risk, I think that it becomes particularly easy to visualize, because I come back to what I am trying to understand is, in risk, is that given these behaviors will occur, how often will they occur or what is the limit at how often they will

occur, and what's the consequences.

So, even when one is setting limits, if there is there is drug interaction between X and Y, and every time it occurs, there is a fatal reaction, then, the 95 percent confidence interval is going to pretty darn high.

But if 0.1 percent is a risk, but it's 0.1 percent risk, and it is a nonfatal risk, and it's a monitorable risk, then, maybe even for a safety concern, that 95 percent confidence interval can prospectively be defined much lower.

So, I can't overemphasize the underdualization based on understanding of the consequence that I think allows an informed decision of where that threshold should be for any of those key messages.

DR. D'AGOSTINO: Again, I agree, and with the confidence interval type presentation, you would be able to show here is the risk that is unacceptable, and the confidence interval is far away it. Here are the risks that are acceptable, and the confidence interval is really tight around

that.

DR. BRASS: Dr. Cantilena.

DR. CANTILENA: If I can just be clear, because I think what you guys were just talking about is important, is that you are talking about the confidence interval approach for the one or two most important items that you want to communicate, and then everything else would then get lumped into a global or an overall, which has historically sort of been in the eye of the beholder. Is that what you are saying, Ralph?

DR. D'AGOSTINO: No, what I was doing actually was in the presentation, I may be saying the same thing, but what I was doing in the presentation is to think of what is really primary, and it could be a set of questions that you have generated a composite, or it could be that you looked at each question separately, but whatever you come up with as being the primary, then, there you display your confidence intervals.

If you have a composite, then, you show your composite makes the little rules we said.

There is the exclusion and then there is the thing you want to rule out and then also the confidence interval, but then I think it is very important that every single variable that is in that composite is displayed for consistency. You don't want it to be just lumped aside. You want to make sure that the ingredients of that composite had consistency with the statement you are making about the composite.

DR. CANTILENA: Okay. Then, just so I can be clear, because I sort of feel like you should not hang your hat on a specific number 80 percent or 85 percent. You are actually talking about what would be a sliding scale, you know, all depending on if it's a safety issue, that is very important versus X, Y, or Z, that is less important.

DR. D'AGOSTINO: Absolutely.

DR. CANTILENA: Okay. I would agree with that.

DR. BRASS: I think that is right, and I would just spin it slightly different, I think, because what I was trying to emphasize, and I think

what Ralph was saying, is that any given study, there is going to be two or three of these primary endpoints that are the most critical issues, which are going to drive the power analysis and are going to have all these prespecified thresholds.

There may very well be a number of secondary endpoints, which can be presented as a composite or as individuals with just exploratory analyses with must lower thresholds, because they are not that important.

In point of fact, this has the secondary effect of focusing the research design and the research effort on those things that matter as opposed to being equivalent between all these potential factors which do not have equivalence in their public health implications.

Dr. Snodgrass.

DR. SNODGRASS: I think this just further emphasizes the need when you don't have prior comparison information, that you need a control group.

DR. BRASS: Well, yes and no. I mean for

example, I think safety is best done on an absolute sense. Just because something else isn't very good doesn't mean that's good enough for the new example. So, I think from a safety perspective, you can say it would be unacceptable if 1,000 people did this, and you have to convince me that 1,000 a year aren't going to do this. I think there is a balance between the comparator and the absolute.

DR. CLYBURN: Just for clarification, I agree with everything that has been said, but with safety issues particularly, it is going to be hierarchical and you have would have to have a higher threshold, if we are going to set a threshold that guarantees standard for switch, like in the question is the FDA going to set that threshold with each predefined objective before the sponsor does the studies?

DR. BRASS: Yes, I think what Ralph has emphasized, and I think what people have agreed with, that for these key points, there is going to be a prespecified threshold that is going to be in

the protocol, in the statistical analysis plan, and it will be in the sponsor's best interest to reach prospective agreement with the agency on as to what that threshold would be appropriate for that specific case.

DR. D'AGOSTINO: I just want to emphasize that it is the sponsor I think that comes forth with what the thresholds are, and it behooves them to get agreement on that.

DR. BRASS: Is that a good discussion of No. 2?

DR. LEONARD-SEGAL: I think so. One of the problems that frequently comes up is picking that number.

DR. BRASS: That's why you get the big bucks.

DR. LEONARD-SEGAL: Again, it's about, you know, what feels good and how stringent we ought to be.

DR. BRASS: Again, I think it becomes much easier for me if you think about it in the context of the consequences of failure, and again, we have

often talked about, well, if this happens, what is the individual health consequence of that.

If the individual health consequence is, as I say, a 0.001 percent event rate, then, you could be a little bit lenient. If it's, as in your recent discussion of Orlostat, if it's you lose your kidney, maybe you need to be a little bit more careful to make sure it doesn't happen.

Again, I don't think any of us could sit here today and give you a number, but I think you are hearing a spirit that that number has to be developed. It has to be developed based on some kind of rigorous clinical context that allows you, then, to make an informed risk-to-benefit decision to the overall population.

DR. CANTILENA: I am not sure I would entirely agree with you, believe it or not, Eric. No, it is just that I think that none of us would be comfortable with a setting where only half of the people got it right, and I think with a critical safety issue, we would be looking at between 80 and 90 percent get it right.

Then, when you get into the subpopulations, that is a hard call, and frequently this committee has seen things that have been all over the place, and you have certainly had to deal with them. I certainly think our job is easier than your job, because ours are just recommendations.

But I think you should perhaps get just a general overview of the fact that none of us I think would be comfortable with a 50-50 kind of--

DR. BRASS: If I said that, I didn't mean to.

DR. CANTILENA: Well, no, but you said that it's all over the place or it's unclear, but I think all of us would be looking for fairly high correct responses.

DR. BRASS: I don't disagree with that, but again the point being that the individual is based on the importance to the individual question, the implications of it. Even in the examples you have chosen, if a sponsor came to you and said here is the prespecified target for this safety concern,

here is the basis for which this target was developed, and, look, we met it, you would have a lot more confidence and a lot more ability to evaluate the reasonableness of that argument than if they simply said it came out this, and that's good enough.

DR. CANTILENA: Which has happened before, so I agree with you now.

DR. BRASS: I think what I would like to do in the interests of time and because we are clearly going to go on a while longer, take a quick break for people who might need it, and we will reconvene promptly at about 2:45.

[Break.]

DR. BRASS: We are going to continue with Question No. 3. In assessing the ability of consumers to self-select, it is often difficult to ask the question without the potential for biasing the answer. Please discuss how self-selection may be ascertained with minimal bias to the consumer.

I am going to go ahead and read No. 4 at the same time because I think it is a little bit

related, and we can discuss them perhaps together for issues of efficiency.

No. 4. Many companies want to use purchase decisions as the metric for assessing self-selection. FDA has refrained from using this metric because there may be other factors that influence the decision which may be totally unrelated to the consumer understanding the label, lack of interest in the product, cost.

How should this type of data be viewed by FDA in the assessment of self-selection?

So, again, these are two aspects, and one of the issues with purchase is its validity of self-selection, so I thought maybe we might be able to discuss those together.

Dr. Fincham.

DR. FINCHAM: Just a comment. There are statistical processes that have been used using different attributes and different levels within those attributes to assess consumer's willingness to do one thing or another.

It is a part of a cost-benefit analysis

approach called "discrete choice experiments," where you give individuals several different things at the same time. It may include price, it may include risk, it may include benefit, it may include cost, and they are nice, tidy, neat. They have been used in the UK from a health economic standpoint for lots of consumer choice issues.

I have used it looking at consumer's choice to use imported drugs, analgesic products whether prescription versus over-the-counter, as well as some compliant behavior, but there are ways to do it within a framework at the same time.

DR. BRASS: Could you really just briefly outline how that flows?

DR. FINCHAM: You give the participant a prefatory series of sentences that assume certain things about an item, and it may be you list what the cost might be, you list what the risks might be versus side effects or lack of efficacy.

You then perhaps can talk about what the benefit might be as far as reduction of symptoms within a certain period of time, and then you just

simply ask the consumer or the participant would you choose to purchase this, yes, no, don't know.

Then, you analyze it with logistic regression techniques, and you can do confidence intervals, you can build in how demographic variables influence whether it's gender or age or whatever, but it does give you a tidy analysis of what would it take to have somebody use/not use based upon these different attributes.

DR. BRASS: Would it be fair to say that the utility of this approach is particularly strong in delineating the influences of discrete, very specific factors in the decision-making?

DR. FINCHAM: You lay out what the specific attributes are, you are correct, but you do give the consumer then a choice to say yes, I would do it, no, I wouldn't do it, and then you analyze it statistically to determine significance. It is just one technique.

DR. BRASS: So that if, for example, there was a concern that an actual use study had the purchase price as a major determinant of the

outcome, one could use this in a parallel study to discriminate that role?

DR. FINCHAM: You could, yes.

DR. BRASS: Other comments?

I just want to say something about the purchase decision. I must admit I have always been biased to liking the purchase decision versus the self-selection yes/no question. The reason, it's completely not data driven like everything else we are talking about, and so it's my bias.

My bias goes like this, that because of what I am really interested in is what is going to happen in the overall marketplace, the only people really who are going to be exposed to harm are those who make the decision to purchase and use the drug, that people might think it's right for them, but not buy it.

From a decision analysis perspective, I think those are two very different thresholds for a consumer participating in a clinical trial, that if you ask them is this right for you, answering "yes" has absolutely no consequences to them if they are

wrong.

So, their willingness to say "yes," I think has a lower threshold than if there is a degree of uncertainty about that assessment.

Whether that would translate into the behavior of actually buying the product and using it seems to me to be a much more robust kind of step on their part that they are not going to take unless there is a higher degree of confidence.

Because I don't care if they think it's right for them, but wouldn't buy it and use it, that has always biased me to want to know what the purchaser's profile is more than those who thought they were using it.

Now, I understand completely that the degree to which the purchase price is an unknown variable in biasing that decision, it introduces a new factor where to the degree the purchase price is a barrier to purchase, but one could always ask the self-selectors who said "yes," who didn't purchase, why didn't you purchase it. If the answer is "price," then, one is concerned.

If the answer is "I just wasn't sure," then, maybe that is other informed, but again I can't overemphasize that is completely subjective, and not based on any data, but it is why I have always kind of thought the purchase decision was the most relevant in assessing the population's safety and the behaviors of consumers who are more likely to use the product in the actual marketplace.

DR. PARKER: When I think about this, and you clarify, if I am wrong, correct me, but it seemed to me that the purpose of the self-selection is for a consumer to be able to show you that they can pick up a label, look at it, and it is my understanding it is supposed to always be on the front. You are supposed to be able to look at the front of this and get this information.

You can make the accurate diagnosis because there is no learned intermediary, that you are using the information on the label to be able to appropriately self-select as someone who could use this medicine. Is that true or not?

DR. BRASS: No. I mean the self-selection decision is integral of all the criteria for self-selection. So, it uses the entire Drug Facts label, and even any supplemental--well, it is only the Drug Facts label--to make the decision whether there is not only the presence of the indication, but the absence of the contraindications based on a personal health assessment.

So, for example, if you have heartburn, but are also vomiting blood, and you select, that is an inappropriate self-selection based on the Drug Facts label.

DR. PARKER: So, it's your ability to self-diagnose that you have the condition and appropriately recognize if you have specific contraindications that are of high enough safety concern. Is that accurate?

So, the questions need to reflect that very specifically.

DR. BRASS: That's correct. So, what is commonly done in a self-selection study is--again, this goes to the bias question--that because you

don't want to cue any of those, one wouldn't, for example, typically, not collect an extensive medical history prior to asking for the self-selection decision.

One wouldn't ask have you ever vomited blood, because that would cue that as a potential factor, so that the self-selection decision is the consumer and the label, and then looking at the consumer's--after that decision has been made--collecting the information to evaluate whether or not that was a label-appropriate decision or not.

DR. PARKER: I guess the way I would think about that would be to say, the same way I take a medical history, I mean I would go back and say I want you to take a look at this and I want you to tell me whether or not it is appropriate for you to take this, tell me why it is, tell me why it isn't. You know it's a series--no?

DR. BRASS: Well, again, you didn't ask the why it is and why it isn't until after they told you whether they would or wouldn't.

DR. PARKER: Yeah, I got that, but it's the level of detail, of not just yes/no, which is a 50 percent chance, but it's the why.

DR. BRASS: Well, again, in terms of scoring the success, it's the correct/incorrect, or it could be acceptable, and what many of us have said, and the agency has suggested, is that if you know the why, you will learn more about your product than if you don't ask the why.

But the threshold for success is simply do patients appropriately select the drug for use.

DR. PARKER: Not based on chance, but based on understanding.

DR. BRASS: Well, hopefully, the threshold for success, according to my colleague, is better than 50-50, so it would be hopefully better than chance.

MS. MAYER: So, I think what you are suggesting is that the purchase decision is reflective of consumer understanding of labeling information? Before you answer, it is my understanding that--I mean I sort of agree with the

question as stated, that there are many other reasons why people pick up drugs over-the-counter. You know, my Aunt Mary recommended this. She said it worked for her arthritic knee pain, so I am going to pick that up and not even read the label.

DR. BRASS: That is exactly my point, that the purchase decision is the definitive assessment as to whether that was appropriate or inappropriate.

MS. MAYER: That is where you have lost me. What do you mean by "appropriate"?

DR. BRASS: That the assessment is made after the purchase decision whether or not the individual's health history as compared to the label instructions is--

MS. MAYER: So, you are saying it's irrelevant essentially, it may or may not include an accurate assessment of whether--

DR. BRASS: No, that is what you are testing, that is the test.

MS. MAYER: But the purchase would not demonstrate that, because there are so many other

reasons.

DR. BRASS: But I am saying if they purchase the product and they were vomiting blood, even if their Aunt Mary told them, that was inappropriate. That would count as a failure.

MS. MAYER: How would you determine that?

DR. BRASS: The same way I did for self-selection. I would ask them after they made the purchase decision, I would collect the medical history. Just like instead of after the self-selection decision, I would collect the medical history.

Question 4 is simply where do you put the check-off, after the self-selection or after the purchase decision, that's the only difference. The information and the evaluation is exactly the same.

MS. MAYER: So, this isn't being collected independently, this is within the context of the medical evaluation you are talking about?

DR. BRASS: I am sorry?

MS. MAYER: I guess I was just reading it a different way, maybe misreading it, that somehow

purchase decisions were some sort of proxy for patients actually understanding whether the drug was appropriate or not.

DR. BRASS: The appropriateness of the purchase decision is an integrated measure of their proper or improper use of the label.

DR. D'AGOSTINO: I just wanted to make sure we don't lose some of the comments that Jack was making. There are procedures for doing this, and I think one of the sort of ways to advise the FDA is that when they are dealing with companies, when companies are putting these studies together, they think of the different dimensions that might be involved in self-selection.

You might be right in terms of the purchases one, but there are different ways, and how are they driving the ultimate self-selection, and when they do finally say "yes," can you look at what happens at these different dimensions. I think that is very important in terms of trying to get information and help on this question.

DR. BRASS: Dr. Griffin.

DR. GRIFFIN: I think it is more conservative not to include the purchase, and I think that it would be really hard to replicate what somebody's incentives for using a drug would be before it was marketed.

I mean you don't have the advertising, you don't have the Aunt Mary using the drug, you don't have your best friend recommending it. So, you might be willing to pay a lot for a drug after you have had all that advertising that you wouldn't premarketing.

To my mind, it seems more conservative to sort of use it as a measure of did I understand it enough to know that this drug would be appropriate for me.

DR. BRASS: How would you react so that if--

DR. GRIFFIN: Because you are trying to conflate two constructs, I think, to say we want to know what is generalizable to what really happens, but I don't see how we can.

DR. BRASS: I don't pretend--I can't

either, I am just trying to get as close as I can, so that if a selection response kind of question scored poorly, but the people who were wrong in self-selecting, ended up being not purchasers, how would that affect your thinking?

DR. GRIFFIN: I think we need to guard against having labeling that a lot of people are going to misinterpret, and that again, if they aren't the people that--that is the safest thing to do.

DR. BRASS: Again, my only concern is that particularly when we get to sophisticated labels, and perhaps some of our earlier suggestions will mitigate against this, that the correctness at that stage becomes relatively complex where the purchase, because of the tangible consequences of its action may anchor that better or it may not, and I agree with you completely actually, that using the self-selection decision is a more conservative, higher threshold kind of decision than using the purchase decision, I actually agree with that.

Neal.

DR. BENOWITZ: I am not sure this has been said or not, it may have been, but it seems to me that both are needed, both self-selection and purchasing, because they measure different things. I am talking about in the actual use. There is no advertising development yet, but still the decision to purchase a drug is based on understanding plus other things.

I think it's those other things. You may concentrate more, you may pay attention more, there may be other factors. I think both things are really important.

DR. BRASS: One of my concerns when there is a bias is that if you--and there is probably not going to be consensus--but if one felt the purchase decision was particularly important, the self-selection questions may bias the purchase decision, and that if you wanted the purchase decision to be the primary and most important, then, it should be the most unbiased, and it, in fact, should be asked first rather than the

self-selection decision.

So, if you wanted to incorporate both, my concern would be that forcing the consumer to make a self-selection decision is part of the cueing about purchase, my bias, the purchase decision.

DR. BENOWITZ: I was thinking more about looking at the data from both the self-selection study and the actual use. The actual use is purchase, and self-selection is understanding.

DR. BRASS: Very often, the self-selection may be built into the actual use, that there may not be a separate self-selection study, but the self-selection decision may be the primary endpoint of the actual use study, and that is where it becomes pertinent.

If you believe my bias, then, self-selection should be--even the isolated self-selection should be replaced by self-purchase, where you don't do the follow-up, but you force the purchase decision rather than the selection decision.

DR. PARKER: But economic access varies

over the life of an individual.

DR. BRASS: In the life of a product.

DR. PARKER: Absolutely, and so I think you can't control for that at the time that you put that--

DR. BRASS: Absolutely correct. Dr. Day.

DR. DAY: We have been talking about the self-selection question as if there is only one way to ask it, yes or no, is this product right for you. You can ask that question and also get a confidence rating, how confident are you, very confident, confident, moderately, unconfident, just guessing.

Then, you change from a two-choice alternative to a four or five, or it could be yes/no/maybe. Then, you have different levels of guessing probability that you can then assess your data in a more sensitive way.

So, it's not just yes/no, and then try to find out why but the degree of certainty that would enrich our understanding about the selection decision.

DR. BRASS: Dr. Snodgrass.

DR. SNODGRASS: I want to introduce an example to just raise some questions. An obese 14-year-old who was distressed over obesity and peer pressure, is now also on cimetidine, goes into the outpatient pharmacy and buys a statin that is over the counter, because it's the anti-fat drug.

DR. BRASS: Yeah?

DR. SNODGRASS: Where does that fit in terms of if you were to design a study to say the purchase decision, there I think reflects a lot of--or they are not going to pay attention to the information.

DR. BRASS: Well, again, if there was an actual use study, and a 14-year-old responded to the advertising and walked into the mock statin display and tried to purchase or self-selected "yes," then, one would obtain the information as to why they thought it was appropriate. This again goes back to understanding the basis for the misinformation or for the misdecision-making whether it be selection or purchase, because that

is your basis for iterative improvement and your basis for assessing whether this is a real problem or not a real problem from a public health perspective, and an individual health perspective.

I would like to hear more about bias about these studies. I have already suggested that obviously, the least information you collect prior to whatever you decide is the primary endpoint, self-selection or purchase, is very important, but does anybody want to expand upon that, or are there issues about how that bias is introduced?

DR. BENOWITZ: There is one source of bias which would be seen if sort of targeted the at-risk population like we talked about. For example, if you want to ask pregnant women if they should self-select, I think they might be better than nonpregnant women, or perhaps, or whatever, chronic renal disease, but that introduced a bias in just collecting those people.

I think you should do it anyway, but just recognize that that may have a bias.

DR. BRASS: I think that's right. I think

it is important and sponsors I think have been pretty good about, when they recruit those patients, to try to make clear not why they are being recruited, that it's not because of their interest in this or that, but they are recruited based only on their medical history, not the purpose of the study has remained obtuse.

But I think what has been less sensitivity to is the diversity of that population, that they often come from single clinics, or they often come from just a few clinics, and they may not represent the same diversity across the consumer population that we expect to see in the larger studies.

I think it is just as important to have that diversity in the special populations to the degree it is relevant, so low literacy renal disease patients, ethnically diverse renal disease patients, socially economically diverse renal failure patients.

I think those issues, again, the groups don't have to be powered based on that, but as Dr. D'Agostino pointed out, looking for consistency in

the point estimates and confidence intervals can be very reassuring that those studies will have some generalizability.

Dr. Griffin.

DR. GRIFFIN: I feel a little hampered by not having seen a lot of these studies before, but it seems like this might be where comparisons make sense, where if you are selecting a group of people who are on coumadin, it might be reasonable to say you have to give them a couple of drug choices, and they would have to choose that a particular drug was not appropriate for them, whereas other were. I don't know if that is the sort of norm of what has been done.

DR. BRASS: It is not, but it's an example of where one can get additional information, I agree.

Dr. D'Agostino.

DR. D'AGOSTINO: I thought I understood the questions, but now the more you ask, the more I am not sure. What is the bias that we are worried about, is that if you ask a lot of questions that

lead them to select, then, you are putting them on one side of the fence or not, is that the--

DR. BRASS: That is correct, the data acquisition methodologies may cue or bias those types of results, is that right, Andrea?

DR. LEONARD-SEGAL: I could respond to this actually. We are sort of interested in the question itself. I think the question itself is the crux of the self-selection study. Medical history has usually been collected afterwards. We have been going in that direction for a while, but it is the nature of the question.

Let me throw out a thought that comes to mind as I am sitting here. We have been talking about multiple-arm studies for these different trials. Do you want to talk about multiple-arm possible studies with regard to self-selection, questions or different pricings for purchase decisions, just for discussion purposes.

DR. D'AGOSTINO: There is where a comparative arm would be helpful because you could compare did you purchase it versus the whole set of

questions that Jack was talking about, and see how the individual--I don't know how you would evaluate what is right and what is wrong, but you could see if the response to the self-selection is different depending on the amount of information and the amount of questions that you gave to the individuals.

Asking the question, do you select, yes or no, the less information in some sense, the better after they read the label, but you could leave that out and that would be a perfect place for a comparative study.

DR. BRASS: Again, just to emphasize the practical aspects, because actual use studies, because of their longitudinal nature are often the most expensive, time consuming, et cetera, that it would seem to me that would only be justified to do a full actual use comparator if there were some critical question that could only be answered through the full actual use process as opposed to a pilot self-selection study, a pilot label comp study, a pilot self-purchase study where the

question is always clear.

I did not feel the groundswell of enthusiasm for the purpose versus self-selection, so I sense that the committee feels like Dr. Griffin that maintaining the conservative threshold of self-selection is something that you are more comfortable with, is that fair? Okay.

That means that in terms of primary versus secondary, self-selection will always then be the primary and the least biased of the outcomes from those kinds of trials.

DR. PARKER: I was just going to add that I feel like self-selection is one of the entities that needs to have tight confidence intervals. I mean this is really important, if you can't accurately self-select--

DR. BRASS: Again, I would disagree, because not all incorrect self-selection decisions are the same. So, take the example of the drug that has got an age limit of 55, and a 54-year-old uses it, and we all agree there is really no public health consequence to that 54-year-old using the

drug.

That is clearly wrong. We clearly, for some reason, wish that hadn't happened, but that is not the same as the renal transplantation on cyclosporin who risks losing their kidney. So, again, I think that one has to be contextual even under such--and I agree, it becomes the critical primary endpoint, but to say that we have got to--depending on the drug and depending on the situation--have that be really, really high may not be in the public health interest.

DR. PARKER: I would just say that given the oversight of the FDA for labels and their content and the lack of an informed intermediary for over the counters, that self-selection is the first marker of being able to engage as a consumer in a relationship with the product, and I would put that as Step 1.

The first thing you do is you have got to pick it up and you have got to look at it, and you have got to decide whether or not it is appropriate for you. I would argue that it does need adequate

power, it needs to be from representative sample, and it needs a tight confidence interval. That would be my opinion anyway.

DR. BRASS: I agree with you two out of three, that's not bad.

Other people like to comment on this because it is a very important question. It's an extremely important question. Dr. Neill.

DR. NEILL: Actually, related to your first comment, I am actually okay accepting purchase decision as a metric so long as the additional data that informs that context can be obtained.

I think in the past where these conversations have come up, I, as a member, have felt hogtied a bit when something about price comes up, and, oh, we can't think about or talk about or something about market pressures come up, we can't think about or talk about.

So if you want your cake, then, let us eat it and give us data about what we anticipate is going to happen in the market. I think that it is

interesting that when we talk about labels specifically--and I couldn't agree with you more, Dr. Parker, people are going to pick it up--but if the purchase decision is part of it, they pick it up off of a shelf where there are a preponderance of labels that are not under the purview of the FDA regulatory oversight, and yet which are made to mimic in order to adopt the brand of the FDA, so there is that kind of distinction, ability to distinguish that the consumer engages in, that also then ought to become a part of that market and purchase decision, not to mention issues of formulary, yada, yada, yada.

DR. BENOWITZ: Could you just clarify, because I am not sure that I have even thought about this, but in an actual use study, so subjects pay for their drugs, or do they get them for free?

DR. BRASS: Yes, no, they pay for them. They actually pay for them. They are usually reimbursed at the end, but they don't know that going in, so they actually have to pay for it.

DR. BENOWITZ: Is there any relationship

between what they pay for it and what the drug ultimately costs?

DR. BRASS: Sponsors uniformly reassure us that the price that is used in actual use studies are representative of their projected marketing plan, but, of course, that has no bearing on second entrance into market, it has no bearing on promotional activities subsequent to release, et cetera.

That is why we always say that, if they are faced with making a decision based on price, you are kind of dead in the water anyway unless we use tools, such as Jack was suggesting.

I don't want to lose Dr. Parker's point here about the standards we should be using here, because I think they are extremely important, so maybe we can incorporate it then into the next question, as well.

Question 5. It can be difficult to verify specific aspects of a self-selection decision. For example, verification of a consultation with a participant's personal doctor can be burdensome.

Under what circumstances is it necessary to verify these components of the self-selection decision and how should verification be accomplished?

Again, because many of the contraindications on a Drug Facts label are relative contraindications, and are under the Ask a Doctor, so even if the patient has the health condition, if, in fact, they say, "I am going to ask my doctor," they get full credit.

Now, that doesn't ask whether or not they actually would. It doesn't ask whether they would communicate the question properly to the physician if they were able to reach them. It doesn't ask if the physician told them what to do, they would follow those instructions.

So, there is kind of a big gap here, and on a number of studies, it is not an uncommon answer. Even the personal health history, if a person says I don't have hypertension, well, what is the specificity of that particular answer on A patient recall?

The question then is in terms of looking at how these studies are done, where is the line between tell us everything, or we trust you, we are doctors.

Dr. Cantilena.

DR. CANTILENA: I think here you really have to look at what are the consequences of not asking the doctor, so it depends. If there are serious health consequences, or, for example, if you are going to be approving a drug, which is now for long-term treatment, and there are some possible consequences for not consulting at some point, I think that would be a case in which we should be able to utilize some schemes to have a look. Perhaps, you know, one of the statin projects had done their studies in the setting of pharmacies that were owned or operated under an HMO, so they were able to go back and actually look in the records to see if a contact had been made.

So, there are ways in which you can do that, but I think, in my mind, the only time you should really consider that is when the

consequences of not asking a physician are significant, or you are going into a new area with the long-term treatment with a drug, that could have consequences if it's inappropriate use.

DR. D'AGOSTINO: I agree that it's important if there are serious consequences, but I don't think there should be a question that says, "Did you have a consultation with your doctor," that you can give an answer that is completely invalid, and if you put that question in the study, I think there should be some mechanism for verifying it.

I do think the consequences can change quite a bit, bit I mean I think the more objective you can be in a study in terms of verifying the responses, the better the study will be, and you presumably will want to make something out of that information when it gets presented here or when it gets presented later on in advertisement, and so forth. I think it is very important if you ask a question like that, that you have a way of verifying it.

DR. BRASS: Since you asked the question, and given the practical considerations, could you comment on the validity of random sampling techniques to verify based on not the entire cohort, but a subset of the cohort?

DR. D'AGOSTINO: I think that is one way of doing it. You don't have to look at every single individual. You can go to the places where they are under the HMO framework and naturally get that information, but the idea would be if you have a 300-subject study, go and get everybody's information. If you have 1,000, then, maybe you can do some sub-sampling, but I don't think that a question like that should be asked and then just left to a completely unverified answer.

DR. BRASS: Dr. Taylor.

DR. TAYLOR: What would be the impact of the HIPAA regulations and verifying that? That would make it very, very cumbersome to do that.

DR. BRASS: Then, don't ask the question if you can't answer it, I agree 100 percent, but you could ask the subject, "Can I approach your

doctor to get that answer," and you can ask within the HMO. I mean there are simple ways of asking that, but they are more and more complicated.

DR. DAVIS: I have a couple of comments about verification. This is when I would agree that people with low literacy may be a different sub-sample, because if you said, "Do you have hypertension," they may have no idea what you are talking about. They say "no" when they know they have high blood pressure.

Also, I want to point out that the elderly have about eight different doctors, so which doctor are you going to call to verify this.

DR. BRASS: The one they said that they spoke to. We do it in Framingham in a lot of these studies. Never mind the low literacy. You go to the nurses, "Do you really trust the blood pressures they send in?" Do you trust the weight measurements they send in?" I mean no.

Other comments on this question?

DR. LEONARD-SEGAL: I guess the discussion is leading towards the fact that verification seems

to be something that might be helpful?

DR. BRASS: Well, again, I think what you are hearing is it depends on the specific context. Where a particular response by the consumer, if the accuracy of that response is absolutely critical to the judgment that the drug was being used safely or was part of the algorithm for safe use, and there was reason to suspect that the consumer's answer may not be accurate, again whether it is a part of the medical history, whether it's contacting a doctor or some other behavior, then, verification is necessary. As Dr. D'Agostino said, if it's not critical and it's not trustworthy, don't bother asking it in the first place.

It seems like for that kind of information, if it's worth asking, it is worth asking in a way that allows you to have confidence in your conclusion. Is that fair, Dr. Parker?

DR. PARKER: Yes. I was just going to add just for sort of specificity, may have an opportunity to garner data on this in pregnant

women. There is a warning for everything about being pregnant, and most pregnant women are under the care of a provider, so that may give you access in one cohort that might be something that is useful later on.

DR. BENOWITZ: I would like to raise the question of what if it's an issue regarded to efficacy, but not to safety, but it's the reason why someone is spending money, and again the statins.

If someone's cholesterol is normal, they are not going to be harmed by taking the statins, but they are going to be spending a lot of money for a lot of years for no reason, and if the criterion for using it is a certain LDL level, then, to me it is important to do that even though it is not a risk.

DR. BRASS: I think that is right, and let me just disclose. I am a consultant to Johnson & Johnson-Merck, but I think that when that was discussed, I think one of the points the committee made, which I thought was important and

interesting, was the emphasis on allowing that consumer to understand the magnitude of their benefit, that it not no benefit, it's small benefit, and how small is that benefit, and can you make a consumer understand that, to allow an informed individual risk to benefit decision whether or not it is idealized or not.

That again goes to this issue of being fully label compliant or within the spirit of the label.

DR. BENOWITZ: But to do that, wouldn't that require someone to have an objective verification of their LDL cholesterol?

DR. BRASS: Which was available in that study. That is a good example. If that study had been done only on the basis or recalled LDL's, one would have much less confidence in making those decisions. That is a good example.

Next question. I think we have already answered this, but I am going to read it, and if there is any additional comments, we can make them.

Question 6. Consumer behavior studies are

generally open label, single-arm studies. Discuss under what circumstances FDA should request that multiple-arm studies be considered whereby the differences in the arms reflect a comparison of different labels or differences in ancillary measures, for example, a package insert versus no package insert.

I think we have discussed that pretty comprehensively. Does anybody have any additional comments about that particular concept?

Just again to summarize, I think there was high enthusiasm for appropriate use of those kind of comparisons to inform individual programs in the general field.

Next question.

Question 7. OTC products may be used intermittently, or have limits on the duration of continuous use, internal analgesics have 10-day limit for pain treatment, or have a set period of use to achieve clinical benefit, the nicotine replacement products.

Please discuss the factors that should be

considered in determining the duration of actual use studies.

Again, I think this is a very individualized kind of situation where one has to think about what are the behaviors that one is worried about over time, and how long is it going to take to see them.

One could imagine a situation with a recurrent disease where symptoms are temporarily relieved, but they recur periodically, and you might care a lot about whether the self-treatment is continued intermittently or as the label recommends, at that point they need to see a doctor.

If that was a concern, one would have to design a study long enough to capture the relapses and the opportunities for repurchase. So, again, I think that if one understands what the key issues are, this is an example where, under Ralph's kind of construct, these become the critical determinants, the critical endpoints in the study, and the duration is them driven by those.

Other thoughts, comments?

DR. D'AGOSTINO: I believe you said it, but in terms of these particular studies like say, for example, the nicotine replacement, I think it is useful to know did the effect happen, and then was there a relapse that the study should build in a long enough period to pick up the sustaining of the effect, but I have nothing else to add than what you said. I think you hit the points.

Each will depend very much on the particular product. The only other thing I wanted to add is that in some of the presentations that were made in the open session, there was the statement made, if I interpreted it correctly, that some of the sort of understanding of how people use the drug and the effects, and so forth, get tossed into Phase IV, I think we should be talking about approving on complete packages, and none of this should talk about be extending into Phase IV.

DR. BRASS: But on the other hand, as an example where if these were critical questions, a Phase IV study to confirm would be--

DR. D'AGOSTINO: Absolutely, and it would be continued, but the approval shouldn't be that somehow or other we have got enough information, we think things are going well, we will give you approval, and then, God forbid, something may go wrong with running the rest of the study.

DR. BENOWITZ: The one area where I think this is actually the most problematic are with chronic medications like with statins or weight reduction drugs where people may take it for many years, and I think those drugs, there really should be trials that should be done for a year, a long enough period of time when you really get a sense of what safety and use patterns will be like.

DR. BRASS: And a year because? A year as opposed to six months, as opposed to two years?

DR. BENOWITZ: I don't have an answer to that.

DR. BRASS: Ralph.

DR. D'AGOSTINO: But we do have in statin trials, for example, how long do we have to wait for the effect to show itself, and then sustaining,

so it is a very good question how long is long enough, but each study, again each product has to be faced individually, but those type of questions should be asked and decent answers should be given.

DR. BRASS: This is again where the totality of data needs to be applied where there may be some Rx studies that address safety concerns, et cetera, and so what needs to be replicated in the OTC trials really had to be OTC center questions that are germane. But I agree that long enough again is the right answer and individualizing what that is.

DR. BENOWITZ: I think a real question here, and I don't know whether we need to address this or not, but it is as much efficacy as safety, whether people will take a chronic medication long enough and well enough to make a difference.

DR. BRASS: That is absolutely right, and there we have trouble again with thresholds given the challenges to doing that in the Rx setting, so that again where we put thresholds becomes very challenging, but I agree completely.

Lou, you have a question?

DR. CANTILENA: Actually, I think you covered it, and I would just say what I was going to say is that it all depends for an Rx to OTC, you have all the information from the Rx, so you wouldn't have to go for five years or one year if you already know the long-term consequences.

DR. BRASS: Ms. Mayer.

MS. MAYER: I was just going to underscore the same point that it makes it particularly crucial to do long-term follow-up of the initial clinical trial, so that we really know what happens to patients on statin after 5 years, 10 years, and so forth. That question doesn't have to be reasked in a over-the-counter setting.

DR. NEILL: Most of the comments have had to do with chronic disease, and I think that it is pertinent because I have not been involved at least since my time on the committee with conversations about what constitutes the OTCness of a condition and whether that has changed, whether or not there is still a learned intermediary that is required,

or whether or not a patient can self-diagnose and monitor.

I continue to have real reservations inasmuch as there is not currently any requirements for the label that instruct the patient with regard to when to discontinue a chronic medication for that kind of condition, which heretofore hasn't been OTC.

When patient ask me, "Doctor, how long am I going to have to take this statin," my answer now is, "Until we know better or until something better comes along," and something better always comes along.

How you put on a label for a patient and then operationalize it is something that I just can't get my head around. Secondarily, I think that that question obviously is real important, because it begs the question of what condition can't possibly be OTC.

I had a patient hospitalized last week who fell while he was jogging. He is 75. He has been on coumadin for AFIB for 20 years. His INR has not

been out of range since I have known him. I never see him. He is the patient who has learned how to control his INR that is the basis for those studies that say that doctors suck, patients are good at it.

Yet, I know that he is not the norm, and yet I could imagine an OTC application for warfarin that says we can show that it is safe and effective for use in the OTC setting, look at this guy. I don't know want a discussion of a day's worth of label comprehension to obviate the need for I think the much more obvious discussion about OTCness of a condition and with respect to labels specifically.

I don't believe that you can convey that information about duration of therapy on a label. The reason that you can't answer 6 months or a year is because we don't know.

DR. BRASS: George.

DR. GOLDSTEIN: It may serve as a useful reminder at this point to point out that the Rx to OTC switches, whether they be statins or anything else, remain under NDA reporting requirements both

for adverse effects, manufacturing, for everything, forever basically, and those reporting requirements encompass a large amount of data that is relevant to the OTC area, as well.

DR. BRASS: Thank you. In the context of this discussion, I want to return to something Dr. Parker raised earlier and make sure that we don't lose sight of this point, and that is again the issue of does an actual use study trump a label comprehension study.

That is, is success in label comprehension a true regulatory requirement, is it a true expectation of this committee, or is it a tool to help guide the design of an actual use study, which becomes the pivotal Stage III equivalent?

I would like to hear some more discussion on it. Lou.

DR. CANTILENA: Historically, it has trumped it. You have seen some pretty bad comprehension studies in actual use and the product has done fine in front of the FDA, so I think I care more about the actual use because that is

closer to how it is actually going to be started in terms of consumer use.

DR. CLYBURN: I think we all have to recognize that people learn differently. They may consult their physician, they may do other things, and I think actual use is much more informative in that result, and I think it does trump.

DR. PARKER: I guess I thought that the label comprehension study was to make sure people understood the label, and I have seen presentations where the actual use study and its data preceded the label comprehension, and I don't understand how you can do a label comprehension study after you do the actual use and expect to have meaningful information or input.

So, I am very confused by that. It just seems to me kind of--are you laughing because this isn't true?

DR. CANTILENA: No, it absolutely is true, and I am laughing because you are exactly right. It has happened and I can assure you that at least from conversations I have had with various

sponsors, that wasn't that intent always, but that's the way it happened.

As a result, I guess we have not come to expect a lot in terms of the comprehension, but ideally, you are exactly right. You learn from the label comprehension, and that should be the label that you use in the actual use. That is really the best way to handle it but I was chuckling because there are a lot of examples where that wasn't the case.

DR. BRASS: This goes exactly to the point where the label comprehension study was done after the fact to make sure the box was checked, and a good study could be shown to the committee because I think we would all agree that a well-designed label comprehension program is in the sponsor's best interest to increase the likelihood of success of an actual use study.

But this also goes to where you set thresholds, so when you are setting thresholds for a label comprehension study, that you may or may not actually care about how aggressive do those

thresholds need to be to inform the sponsor and the agency that doing an actual use study is an appropriate thing.

DR. PARKER: I would just take that all the way out to say the recommendation is that the label comprehension study and its results inform the actual use study. This is a recommendation, this is an expectation, and there are actually consequences for that not happening.

DR. BRASS: But why should there be? If they don't do that, and they come in with a perfect actual use study that was done first, and a label comprehension that was done after the actual use using the same label that confirmed its utility, why should that matter?

DR. PARKER: Because I think we can take that actual use and we can re-spin the data, and we can look at it, and we can probably look at it very critically. If every piece of what you suggested in terms of the standards went forth, the power, the carefulness of every single component was there, and I would also argue that if indeed we

have the health of the public in our palm, as our mission, we are trying to do everything we can to help people understand what they need to do in order to make good decisions about their health, and if we don't do it, who is?

DR. BRASS: I agree completely, but given what we know about the predictive value of these various--or we don't know about the predictive value about these clinical trials, I am a little concerned that taking absolutist positions is self-serving as opposed to serving the public health.

DR. PARKER: I would say that that depends upon the quality of the studies and the data that we get from them.

DR. BRASS: Dr. Neill.

DR. NEILL: I think that the fact that these comprehension studies sometimes come after actual use and form certainly my reason for why I believe, as I think Dr. Wood does, that labels don't matter much. A and B, why that actual use study targeted at meaningful outcomes may be

helpful. I would absolutely grant that it can be acceptable to understand why you have succeeded, after you have succeeded, so long as success is defined within that context of public health, and I think that that has been seen.

I think that to the extent that self-selection studies that use purchase, and that actual use studies that mimic as closely as possible that naturalistic environment inclusive of not just standing in a pharmacy or in an HMO, but include all of that sort of rich cultural milieu that our patients walk in, would certainly make me more confident that the actual market effect after the fact will have been reflected by that actual use study, and makes me, as a committee member, more confident saying yeah, this is where it is going to go.

We have talked about these Rx to OTC switches up to now. The reason that I think many of us still feel that those prior studies have been incomplete is because they haven't been predictive of that.

We have seen things happen in the market that have arisen out of the complex environment that the market is, and we need to be able to get closer to mimicking that, so that we can make informed decisions about public health.

DR. BRASS: Neal.

DR. BENOWITZ: The one area that I would be concerned about would be the subgroups where either they might not be enrolled in an actual use study in a middle-class shopping mall, or they wouldn't be appropriate, so I think low literacy people, people with chronic renal disease, the subpopulations, it would be nice to know how they understood the label even though they wouldn't likely be involved in actual use studies.

DR. BRASS: I agree with that. I think that is an excellent point, and that is also where the target of self-selection studies are also I think really important, because of the under-representation in the actual use study where those questions become really important.

Dr. Griffin.

DR. GRIFFIN: I guess I am coming down on feeling like there should be standards for comprehension. Usually, multiple things have to go wrong for somebody to have a medication error. It is usually not just one thing. So, we want to minimize all the possible things, and I think label comprehension is just one part, and it seems like there is a whole science that has been devoted to how people understand things and that we should incorporate some of those, what we know from the science, and that should be the standard.

DR. BRASS: Dr. Day.

DR. DAY: Are there any studies where there is a little bit of label comprehension testing at the beginning of an actual study and at the end to see if the experience in using a drug changes knowledge? It could go up and it could go down.

DR. BRASS: We are talking about label comping done afterwards. It's an independent population, it's not the same population.

DR. DAY: I do understand that, but there

would be some utility in doing this.

DR. BRASS: To my knowledge, I know of no data that tests the actual use cohort on a label comp study.

DR. DAY: It would be interesting to think about what the advantages would be in doing this.

DR. BRASS: Again, you could only do it afterwards for bias reasons, and even then you would be biasing the label comp study.

DR. DAY: But you would be studying something else, so if you have the package with you, do you then consult it more and learn better, or once you use it and decide to use it, you put it away, and whatever knowledge you had decreased. So, I think it's a separate question, but one worth asking.

DR. BRASS: Ralph.

DR. D'AGOSTINO: If you do take the posture that the actual use study can trump all the preceding studies, you do want something in that actual use study to give you comfort that they read the label and understood it somewhat, so how you

build that in is an interesting question, but I think it is a very relevant question.

DR. BRASS: Dr. Clyburn.

DR. CANTILENA: I go back to something Dr. Parker said this morning, and I think that part of the problem with the difference between that is the lack of identifying critical issues in the label comprehension, because I think that actual use study, it makes the label comprehension study less important if they don't hit important issues with it.

DR. BRASS: I am going to go on to the next question.

How should we determine which information is essential for self-selection and use and therefore must be on the Drugs Facts label and what information could be provided in a package insert?

Again, I think we have talked about this indirectly all through the day, and I think there is regulatory and philosophical consensus that the Drug Facts label must contain all the information that is required to make that initial

self-selection material, and that other material is really supplementary.

The way this question is worded is which information is essential for self-selection, well, that goes back to the point that has been made in various ways where that is individualized, and now the individual drug, what are the 3, 4, 6 things that are most important, or in the case of some drugs, 32 things that are most important for ensuring that self-selection is being properly done, and I think that is very individualized.

But I think all these discussions about trial design in this question very much emphasized the responsibility to choose wisely, that when we say something is essential, when we say something is important, it is not just kind of important, it is not just wouldn't it be nice if, that it is really the high-value targets that need to be there for individual and public health reasons, and those are the things we need to test.

I think that is the theme that I have heard repeatedly through the course of the day.

Lou.

DR. CANTILENA: For me, it is who is the drug for, who is it not for, how do you use it, and what are the side effects. Those are sort of the core things for all, I think.

DR. BRASS: But even again for the side effects, as we have heard from our speakers, listing all the side effects has no utility.

DR. CANTILENA: Totally agree. You have to rank/order those.

DR. BRASS: Moving on to Data Analysis and Interpretation. Some products may have multiple criteria for a consumer to consider when determining whether they are eligible to use the product, the ever-present example of cholesterol-lowering.

What standard should be applied when interpreting self-selection data for these types of products?

I think we have already discussed this, as well, when we talked about setting thresholds based on individual criteria, based on their relative

importance, and on using composites for some others, but really trying to individualize this.

Other comments? Dr. Parker.

DR. PARKER: Only just sort of that concept that I think it really is a public health issue to think about not just whether or not it's safe to self-select, but whether or not it was appropriate to self-select from an economic standpoint of someone with limited means who self-selects, and it may not have adverse health outcomes per se, but it certainly takes away from the money that you have available to spend on other things.

So, I think there is a message in there that is actually sort of one of equity and one that we need to pay attention to.

DR. BRASS: Neal.

DR. BENOWITZ: I think this interface is a little bit with Question 3 below. I am personally skeptical that a low literacy person would ever be able to understand the cholesterol guidelines, but that is not to say that the drug shouldn't be

available to people who can understand it.

So, I think you have to kind of consider it in that context, as well.

DR. BRASS: It goes to my Japanese label example, that it is not just whether or not they can, but if they don't, what do they do, because as I say, I think there is a big difference between I don't understand this label, I am never going to spend $30 for this, or I hear this is really good, I have no idea what it is for, but I am going to buy it and take it, and expose myself to risk.

Dr. Neill.

DR. NEILL: This is one of those questions where I think getting qualitative data about the nature of the entire spectrum of patients that present is going to be useful.

At one end there is the patient that doesn't read Japanese. At the other end is the patient that is 75 and fell and on coumadin, and going to do anything without me, and in between there is that vast mass, and I would like to be informed what it takes to move people to this end,

whether or not that is available in their environment, what do we know about what it takes to move people to that end, and what are the consequences of having an agent available in the OTC market without that movement taking place.

I have been impressed that there have been public education projects, and I know that when some entities have come before the committee, they have come not just with label comprehension self-selection, but with the "And here is the education campaign that we are going to do. It is not required, but we are going to do it anyway."

Even that doesn't inform that important question about that, the qualitative data about how to get people sort of to where they need to be.

DR. BRASS: Dr. Parker, I would just like to follow up with the dialogue we had a little earlier about self-selection, and in particular, whether or not you are at least comfortable with the concept that not all self-selection criteria should be evaluated quantitatively the same.

DR. PARKER: Specifically?

DR. BRASS: Again, I will tell you the example, a label for a 55-year-old that is used by a 54-year-old. Should the quantitative criteria for success for that self-selection criterion be the same as one that exposes the user to a substantial individual health risk as the two extremes?

DR. PARKER: I think it answers itself. I totally agree with you, I mean a 54-year-old versus, you know. On the other hand, I think that you should be able--actually, I think I would take a stand on this.

I think that if the purpose--and this gets back to the purpose of self-selection, and this is why I said this the way I did--if the purpose of the self-selection, and you jumped in when I started to say this, relates to the ability to not have an informed intermediary in the process, and that gets into product liability, I mean that is what is under it, if I am right on this.

Now, I may be wrong, but my understanding was you should be able to take the information and

understand it, and be able to demonstrate that you understand it in order to be able to make an appropriate decision.

Now, if it says, "Do not take this medication if you are 54 or older," and I am 55, and you say, "I want you to take this, I want you to look at it, and I want you to tell me whether or not, based on this information, you can take it."

And I know I am 55, and it says 54, and I said, "I can take it," and you say, "Tell me why," and I say, "because right here it says that you have to be 54 or younger, and I am 55, and that means it's okay for me," and I demonstrate clear lack of understanding, I have got a problem with that.

DR. BRASS: What if I say, you know, "I turned 55 yesterday, and I know I have this condition, and I know chronological age is much less important than physiologic age, and I am in really good condition, so I use it."

DR. PARKER: Well, I would say as an investigator or as a researcher, my job is to make

sure the participants in a self-selection study understand that I am there to make sure that this label adequately informs in a way that they can understand, and my study goal is actually to make sure that they can--you know, I have got to have a gold standard, I have got to have a yes or no.

DR. BRASS: I understand that, but to say that those two are equivalent, because again you end up with Dr. Shiffman's and Dr. D'Agostino's variation of when, on a typical label, there are 12 components, and you do each at 90 percent, and then you cumulative correct is going to be substantially lower, and trying to put a 95 percent threshold as a requirement for 12 decisions, some of which matter and some of which don't, is putting a very substantial hurdle.

DR. PARKER: Oh, it's hard, but I would argue that we can do it if we do good studies and good research, and we garner the data, and we put the comparisons in place, and we do some of the hard work it is going to take to figure out how to do it.

DR. BRASS: I actually think it is next to impossible, but we can see.

DR. PARKER: But you take that priority of what information is absolutely critical and which ones are we going to hold to this standard.

DR. BRASS: Dr. Day.

DR. DAY: When we have multiple criteria for self-selection, you pick, say, three of them that are really important whether it's gender, age, LDL level, family history, and so on, take the ones that are most important, do a factorial design. What that means is you cross each one with every other one to generate the problem space.

So, for example, you have Okay or Not Okay on each of the variables. You then generate the whole problem space, and you have lots of different scenarios. Now, maybe it's too much for one person to answer all of them, but across different subsets of people, you can then find out the power of the individual factors.

So, it may be that when there is only one of the factors that is not correct, they may accept

that it's okay for Sam or Judy or the scenario person to do it, but there might be some of the criteria that once it is a No, it's no correct, it is going to override everything else.

I think there is a tremendous richness in understanding that we could figure out if we would use that kind of approach, rather is it okay for you or this person when it is only a subset of these scenarios.

DR. NEILL: I do think this is a setting where the comparison group being the Rx group is useful. I mean we are talking about the Rx current situation as if it's perfect, and I am going to tell you we doctors stink at selecting these patients, and I think that is a real public health concern.

If you have people seeing doctors in a restricted formulary environment where patients who otherwise ought to be selected aren't, that is an issue. So, in some respects, that sets the bar a little bit low for the comparison group and ought to be able to be met in some actual use studies,

but that is just one way in which looking at the whole lay of the land, people who are unable to self-select are able to self-select prescription, OTC environment, who is appropriately being selected in either, would at least help inform that approval decision at the outset.

DR. BRASS: Neal.

DR. BENOWITZ: It seems to me that there are two different questions that are being talked about. One is to have someone, should a person like myself take this drug according to these criteria, and then the second one is would I like to buy this drug. The second will allow you some leeway, to say, well, I know I don't meet this criteria, but I am 54, and therefore it's fine.

There are really two different issues. One is understanding and one is making a judgment, and maybe they should be separate.

DR. BRASS: That gets us back to which is the primary and which is more informative in assessing the marketplace risk to benefit equation, but the other point I didn't belabor with Dr.

Parker, but how the self-selection question is asked is actually more behavioral than understanding.

It doesn't ask if you understand this label. It asks what you are actually going to do - do you think this drug is right for you today, not did you understand the label, and that is where the complementary information from the label comp and the actual use comes in.

I need just a little bit more discussion about the multiple comparison thing, because I have heard some disparities here, and I really think this is of central importance. I really would like to get some broader.

Dr. Clyburn.

DR. CLYBURN: I was just going to say clearly, some mistakes are catastrophic and some mistakes are trivial, and I think they have to be hierarchical in their response.

Going back to some of the examples from earlier, if the reason for the 55-year-old limit is childbirth potential and the person has had a

hysterectomy, they are actually correct even though the tests would say they are incorrect in doing it. So, I think that we do have to look at individual things, and it is not an on and off phenomenon.

DR. NEILL: It is also true that in some of these subgroups, low literacy, for example, the competing comorbidities in that group, given that their mortality and morbidity is going to be higher as a whole, just simply accruing from their low literacy, means that you are asking for a benefit to accrue by (a) appropriately self-selecting, and that it is specifically because they have selected this medication. Potentially, you may think this is a group that is going to benefit all the more because look at how poor their outcomes are to begin with, and yet it is exactly why they might not benefit.

In other words, that patient with the hysterectomy is not going to self-select for the statin if their hysterectomy was for metastatic endometrial CA, and if they do appropriate self-select, given all the other things, then,

would we say they selected appropriately if their life expectancy was six months? No.

If somebody has low literacy, and low literacy is predictive of poor health outcomes, in the absence of data that suggest it is appropriate self-selection for this one medicine and proves that, this is the pill to fix the problem that you can't read, in the absence of that data, why would we presume that there would be meaningful outcome improvement whether morbidity or mortality especially for chronic illness. For short-term things, fine, but for long term, that would be a much more difficult thing to show, I think, in that subgroup.

DR. BRASS: Because of the overlap and because of the hour, I am going to go ahead and read the next question because I do think that they begin to complement each other.

Companies often want to include responses as being correct, even though they do not conform exactly to the labeled information. How should these types of responses be evaluated in the

assessment of consumer behavior? If they are going to be permitted, should they be pre-specified in the protocol of the study?

I will take Ralph's response quickly, because he will say, of course, it should be pre-specified if it is going to count.

DR. D'AGOSTINO: It absolutely should be pre-specified, but, you know, this gets us in a situation where the label comprehension study may not have been put together in a perfect fashion, and you learn from it, and it feeds into the following studies.

But when you come down to the study where you are going to finally do the evaluation and say that this incorrect behavior somehow or other is correct, that all has to be pre-specified before you can really build it into the analysis, and I would say even in the label comprehension, they should be pre-specified. They may find in their analysis of the pre-specify that they wish they had done things a little bit different. They can present that as a secondary analysis, but the again

down the road with the self-selection and the actual use, they have to be very specific.

DR. BRASS: I think it also affects thresholds where you are going to have both correct and acceptable, you might anticipate a threshold that is higher for being good enough than if you are just being correct. I meant there should be interaction there.

I would be interested to hear from the behavioral scientists about whether a label comprehension study could be designed where acceptable was not in the algorithm, where the questions were adequately designed, so that they were correct or incorrect, and we wouldn't need to resort to this.

Dr. Day.

DR. DAY: When we do experiments on anything, we get an accuracy level, we always do a complete error analysis, and this can be totally impartial with respect to approvability or not in this case.

So, the first thing you look at when there

is an error, was there a no response response, and if they did have no response, do they say I don't know, or just--when they have no response, it could be because they don't know, or it could be they don't understand the question. We saw examples of that in the video today.

So, one type of response is a no response response. The other type can be categorized according to a priori criteria, and then after that you can assess what's includable or not, but I would expect, as a reviewer, to see a complete error analysis of all data on all questions to see the range of type of errors that are being committed, and you can often learn more from the errors than the accuracy levels.

DR. BRASS: Let me give you another concrete example. On a label comprehension study that is multiple choice, should "Ask a doctor" ever be an option? I mean it seems to me the answer to that is no, because it is always going to be right. How could that ever be wrong?

So, if that is not the correct answer, it

becomes acceptable, and then just increases the noise around which you are really trying to test. I mean that is the kind of specifics I was trying to get at.

DR. PARKER: Or circle all that apply.

DR. BRASS: You mean expecting them to circle more than one.

DR. PARKER: Yes.

DR. DAY: In the open questions, that is a particular place where you can categorize all responses, but you are absolutely right, when there are response alternatives given, the onus of the burden lies on the designer, and you can load it, and please don't forget what guessing probabilities are. You know, when you have got a yes or no, it's 50-50, but when there are four responses, it is 0.25, and so on, so it is not just total percent correct, but we have to adjust for the guessing probability.

DR. BRASS: And particularly when there are four choices, one of which is correct, two of which are acceptable.

DR. DAY: And also how bad are the other ones? I mean are the foils, the ones that aren't right, really outrageous, and no one would think of, or are they plausible, so the plausibility of the foils.

DR. BRASS: Continuing.

How should data from low literate subjects be evaluated relative to data from the general population of subjects included in the studies? Alternatively, should FDA just require a certain percentage of low literate subjects be included in the study and conduct analysis only on the whole population?

So, again, this is something we have talked a lot about. Ralph.

DR. D'AGOSTINO: You could have a focus in the study on the low literacy population and design the study so it talks about the overall and talks about this group also. That is perfectly acceptable.

You can ask them to have some kind of computation for the number of low literacy and then

do an overall analysis, but I think you must do a subset analysis, and the analysis should consist of your overall group and then low literacy, higher literacy, male, female, ages, and so forth. I don't see that in any way that you should leave the analysis solely sort of blind to what happens on this low literacy group.

It should be looked at and should be consistent with the overall results.

MS. MAYER: Let me just clarify, how are you defining low literacy?

DR. LEONARD-SEGAL: Our labels are targeted to 8th grade. I mean that is I guess a debatable question, as well. We are under the impression that it would be almost impossible to write these labels for lower literacy targets.

DR. PARKER: I don't know if that is true. I don't think that has been proven. I think that that would be an assumption. I don't think it is true. I think if the incentives are there, it probably could happen. I don't think it has happened yet.

Let me just say this another way. The average reading, which is decoding skill of an American, is at the 8th grade level. That is the average. Comprehension is about two grades lower. That is for the population at large. So, when you are targeting the average American under mandate, you are targeting a reading at 8th grade, a comprehension at 6th grade. That is not low literate, that's average unless you want to say everybody is low literate.

I think just to be very clear in what we mean by that, you know, we want to norm it at the right place. I just want to be clear on that.

DR. BRASS: I think that is very important.

I want to try another way to ask this question that I have raised a couple times. What would be the committee's reaction to a drug whose label was relatively complex, and it was just clear that because of the properties of this drug, it just was going to be complicated, and that is the only way the label was going to be.

In fact, the average consumer could not comprehend this label. But in a self-selection study, those who couldn't comprehend simply didn't select, and those who selected, selected appropriately.

What would people think about that from a regulatory and public health perspective?

DR. PARKER: That you violated the intent of the law.

DR. BRASS: I am empowering you to make a recommendation within that spirit, because, in fact, the people who selected, the decisions being made were not inappropriate. If you didn't understand the label, you didn't use it in the hypothetical.

Ralph.

DR. D'AGOSTINO: Can you imagine three months after it is approved, the television commercials will tell you it is available and appropriate for everyone? I don't see how we could really do that although I think a number of my colleagues would say that they are as smart as the

M.D. and they should be able to select for themselves.

DR. SNODGRASS: It seems to me that some of that group that you are saying wouldn't select it, actually, probably would select it without understanding.

DR. BRASS: Suppose in the extreme hypothetical, because if the extreme hypothetical isn't acceptable, in reality--because again this has corollaries to me about how you analyze the low literate and how many is even enough to include in the study, because what you are saying is you need to have confidence that the low literate group will be able to use the drug safely and effectively, using the standards for safe and effective that were prospectively identified.

Now, again, that could be done as part of the overall group with representation and just show that the low literacy was not way out of the bounds on a panel kind of display as you indicated previously, but you know they are going to be worse. They always are worse.

So, are we really saying that the low literacy group by itself needs to meet those thresholds, is that really what you are saying?

DR. D'AGOSTINO: We have mentioned that over and over again. I think that if the label is pegged at low literacy comprehension, and there is a consistency with that group, as the subgroup, that that is sufficient.

DR. BRASS: The final question I think we have also discussed previously.

What type of information can provide more confidence that these studies are predictive of actual consumer behavior in the marketplace?

That is more research about consumer behavior in the marketplace and Phase IV studies that look at key issues to ensure that the behaviors predicted are, in fact, occurring, would those raise substantial safety concerns for the public health.

Any other comments? Any other overall comments?

DR. PARKER: I just had one comment. I

wanted to just thank you all for putting this on the program for us to go over. Obviously, I care a lot about it and I think it is very important, but I would like to think that perhaps it is something that we revisit on a regular interval given the agenda for improvement, and that we not look at it as a once every 10 years kind of thing, but we really try to take a look at where we are in a fairly short term and sort of what our goals are for overall improvement.

I think we have got to roll a ball up hill, but I think it can be done.

DR. BRASS: Again, hopefully, you will see that as each new NDA comes before the committee.

Charlie, Andrea, anything that we haven't touched on or anything that you would like to say in closing?

DR. GANLEY: I think one of the things that Andrea mentioned to me earlier that raised a little concern is that in setting these thresholds, if we decide that there is a certain threshold that has to be met, the last thing we want is to come to

the committee and have them say what the heck were you thinking when you said this.

I guess it would be helpful to get some sense as to what are the percentages we are talking about. I think it is easy for the things where, if they are wrong, the consequence is not that relevant, but what if the consequence is relevant, are we talking that the lower threshold shouldn't be below 90 percent, you know, if you do it by a confidence interval, or what are we talking about here?

DR. BRASS: First of all, my first answer to you is to mirror what you always say to us. It's the discussion that matters, not the answer. In fact, if you came to us with a threshold, you, sponsor, came to us with a threshold and articulated the basis by which it was derived, I think that would have considerable influence and mitigate some of the negative visceral reaction, because I think that again I think there will be some thresholds that are over 90 percent, in targeted subpopulations on critical issues, and

others will be more lenient, not 50-50, but maybe 80 in terms of what is good enough.

I think that, not personally to dodge the question, but it is so individualized that what I think myself and others have articulated is a rational process that individualizes the threshold determination based on the individual and public health consequences of not heeding.

Ralph.

DR. D'AGOSTINO: You would like also that the question has been thought of before the data analysis. I think that I have seen before, again in my limited experience, is you have an analysis of the data and then you look at a rate that 80 percent said it correctly, and then you say, well, is 80 percent good enough, and you are scratching your head on how to figure it out.

What I think we are asking for is that you think out what you expect before you actually put the study together or before you run the study.

I think some of the things, you know, with the safety, for example, I mean again if you think

of how we put safety studies together in say, Phase IV prescription drugs, we know what the background rates are. You can get a sense of is this drug, if you are within 1 percent or a relative risk of 1.5 of the background rate, is there going to be serious consequences, and so forth, so there are many situations where you can put the other reasonable number.

DR. LEONARD-SEGAL: I think that again we will be visiting this stuff and I am wondering whether or not we might be considering bringing threshold questions to you ahead of time as we are discussing this process.

I am just throwing this out in the air now because it becomes a potential sticking point.

DR. BRASS: Again, I think that would be quite possible except I can't imagine it being done in a way that doesn't delay sponsor's program substantially. There is a parallel to this in other divisions, the threshold for non-inferiority trials, the thresholds for relative risk for safety, and informed judgments are made, and

advisory committees, and particularly in this setting because as Ralph and others have emphasized, that compared to the post-hoc arbitrary prospective--it is such a quantum leap in what we are doing that I think that the likelihood that you would be laughed out of court is kind of low.

So, we trust you. I am more afraid of giving you a specific threshold that is inappropriately high or inappropriately low for a specific question and have that come back and get thrown around. We met this threshold and that is what they said at the meeting, why are you giving us a different threshold when they said 0.85 at the meeting.

DR. CANTILENA: I think you should consider generating some sort of a guideline that has a little room for flexibility, but a guideline just to get folks in the range, and especially as Ralph has probably said 100 times today, it has to be set out in advance a priori, and I think if you do that, just like Eric just said, that would be a huge help to everybody.

DR. BRASS: If there are no other comments, I would like to thank all the participants particularly those who gave us such informative presentations over the course of the day. I would like to thank Darrell for his help in organizing and getting everything set up, and thank you all for your attention.

We are now adjourned.

[Whereupon, at 4:15 p.m., the meeting was adjourned.]

- - -