FOOD AND DRUG ADMINISTRATION

INSTI

TUTE FOR SAFE MEDICATION PRACTICES

PHARMACEUTICAL RESEARCH AND MANUFACTURERS ASSOCIATION

* * *

EVALUATING DRUG NAMES FOR SIMILARITIES:

METHODS AND APPROACHES

* * *

PUBLIC MEETING

* * *

THURSDAY,

JUNE 26, 2003

* * *

The meeting was held at 8:30 a.m. in the Grand Ballroom South of the Renaissance Hotel, 999 9th Street, N.W., Washington, D.C.

PRESENT:

MICHAEL R. COHEN, M.S, Sc.D., Institute for Safe Medication Practices

ROBERT E. LEE, Jr., J.D., Eli Lilly

PRESENT (Continued):

CAPT. THOMAS G. PHILLIPS, Food and Drug Administration

BESTON JACK ABRAMS, ACT, Inc

WILLIAM H. CAMPBELL, Ph.D., University of North Carolina at Chapel Hill

SUZANNE COFFMAN, Pharm.D., NDC Health

JAMES L. DETTORE, Brand Institute

SHARI DIAMOND, J.D., Northwestern University School of Law

BONNIE DORR, Ph.D., University of Maryland

CLEMENT J. GALLUCCIO, Interbrand Wood

JOHN GOSBEE, M.D., P.E., Veterans Health Administration

PETER A. GROSS, M.D., Hackensack University Medical School

THOMAS H. HASSALL, M.S., Merck

KAZ JASZCZAK, Parascript

JOHN K. JENKINS, M.D., CDER, FDA

BRUCE L. LAMBERT, Ph.D., University of Illinois at

Chicago

TIMOTHY LESAR, Pharm.D., Albany Medical Center

PRESENT (Continued):

SHARON OLMSTEAD, Pfizer

SUSAN PROULX, Pharm.D., Med-E.R.R.S.

PAUL S. SELIGMAN, M.D., CDER, FDA

R.F. SHANGRAW, Jr., Ph.D., Project Performance Corporation

TONI M. STIFANO, CBER, FDA

BRIAN L. STROM, M.D., M.P.H., University of Pennsylvania School of Medicine

MAURY M. TEPPER, III, Womble Carlyle Sandbridge and Rice

KASEY THOMPSON, Pharm.D., American Society of Health-Systems Pharmacists

SUSAN C. WINCKLER, R.Ph., J.D., American Pharmacists Association

DAVID R. WOOD, Interbrand Wood

C O N T E N T S

C O N T E N T S (Continued)

PAGE

P R O C E E D I N G S

(8:01 a.m.)

DR. COHEN: Good morning, everybody. Could everybody please be seated?

Welcome. Thank you very much for coming.

We're on a tight time frame. So we're going to try to stick to it to the minute actually.

My name is Michael Cohen. I'm from the Institute --

PARTICIPANT: That microphone is not working.

DR. COHEN: Thank you.

I was saying we're on kind of a tight time frame. So we're going to try to stay to the minute actually today. We have a number of speakers also as you know and also a public comment section.

I wanted to welcome everyone and thank you for coming. Thank you for your interest in this subject, and I'd also like to thank my colleagues who work with myself and others to put this meeting

together:

Bob Lee from Eli Lilly Company, a trademark attorney with PhRMA, the trademark attorneys in PhRMA.

Jerry Phillips from the Office of Drug Safety, who is directing right now the Division of Medication Errors and Technical Support within ODS.

Mary Gross from FDA, who kept us on track and got us together a number of times for teleconferences to design this meeting and just was instrumental in pulling it all off.

Thank you, Mary.

And Allen Vaida, a colleague of mine, who's Executive Director at the Institute for Safe Medication Practices. I thank you as well.

I obviously always would want to thank all of the participants in this meeting, the speakers and the experts that we've invited to participate.

I think this is a really good news story for all of us, as a matter of fact. I know that you know the Institute of Medicine published a report in

1999 about the problem of medication errors, and actually within that report, there was quite a bit of discussion about problems with drug nomenclature that were occasionally leading to medication errors.

And I can tell you working with the USP-ISMP Medication Error Reporting Program and also I'm sure Jerry would tell you with the FDA's MedWatch Program, a large number of the errors that we get reported from the field that affect our patients have to do with name confusion of one sort or another, not just brand name, but also nonproprietary name, abbreviations, et cetera, et cetera.

But the good news is, in fact, this has been recognized for some time. Really it goes back about at least 12 to 15 years ago when FDA became very interested in this subject after reviewing reports and put together some groups within the agency to look at NDAs and look at the names that were being proposed.

And as time went on, other organizations got involved with this, and many companies hearing

from practitioners from around the country made the decision on their own to initiate the testing of brand names, the proprietary names of their pharmaceuticals to try and help assure that there would not be confusion with that medication.

And I think to a large extent there has been a great deal of success in that area. As a matter of fact, I think anyone that works in this field would be able to tell you that many names that might have been problematic have been kept off the market with the current system.

Unfortunately, the fact of the matter is that we do still occasionally see drugs marketed today, and although they are tested by various consulting companies and tested by the companies themselves in many cases, certainly by people within the agency, we do occasionally still have drugs that reach the market, and then once in practice people begin to prescribe these medications, list them on computers, et cetera. We begin to hear that, in fact, there has been a mix-up.

So this good news that I talk about is not criticizing what's been done in the past at all, but trying to improve it, trying to figure ways of improving the system, the methods that are being used to help to better assure that we won't have mishaps with medication once the product is launched, and I think everyone that participates here today understands that, and that's what this is all about.

We really want to do the best job possible, not knowing what that is at this point, and so that's why we invited various scientists from this field from around the country to participate in this. They've made themselves known to us over the years through their work, and we've invited them to comment in various areas that you see in the program.

And so I do believe this is a good news day today, and I think we'll all walk away at the end of the day feeling quite a bit better than we did when we walked in.

So with that, I again would like to thank everyone who helped to put this meeting together, and

now I'd like to call on Bob Lee from Eli Lilly Company and PhRMA Trademarks to say a few words as well.

MR. LEE: Thank you, Mike.

On behalf of PhRMA I'd like to welcome everybody to the meeting, and I think it's going to be an historic meeting. It should be of interest to many different parties.

Being early on the program, there's a certain advantage. I can be the first one to mention today the FDA new acronym, MEPA, M-E-P-A. There's a number of ways you could pronounce it, but I like to say MEPA because it emphasizes the "me," which is the individual efforts that I think everybody has to bring to bear to try to solve problem. MEPA stands for Medication Error Prevention Analysis.

The focus today is on trademarks, and I think with the panels that we have and the experts we have here today that it's going to be a very successful meeting. So in the interest of time and to get on with the rest of the program, I'd like to

introduce Captain Jerry Phillips from the FDA.

CAPT. PHILLIPS: Thanks, Bob.

It's a pleasure to be here, and we are very pleased to be a partner here with PhRMA and ISMP for this meeting. We have all worked very hard to be here today to put together an opening dialogue. This is the first of a dialogue to discuss the methodologies on how we test trade names, and with that in mind, some of the purposes of the meeting was to look at the current processes that we all undergo both in our companies, at the agency, and at private companies, and then to have the perspectives of those particular companies; also look at and have experts, independent experts come up and talk about the different methods so that we can have a discussion and a dialogue. There will be an opportunity for the public discussion during the meeting.

And so with that I think we're really excited to be here today and have an open and friendly discussion about this very important subject matter that will improve patient safety.

And with that, I will ask Paul Seligman to come up. He'll give us an overview of our expectations today.

Thanks.

DR. SELIGMAN: Let me bring up the rest of the panel as well for this morning. Dr. Jenkins. Is it Sharon Olmstead? Tom Hassall, are you here? Tom is here. Good. And Dr. Lesar, come on up and join us.

While people are taking their seats, let me wish you a good morning. My name is Paul Seligman. I'm the Director of the Office of Pharmacoepidemiology and Statistical Science in the Center for Drug Evaluation and Research.

It's a pleasure to be here this morning to welcome you to this FDA public meeting that is being co-sponsored by the Institute for Safe Medication Practices and the Pharmaceutical Research and Manufacturers Association.

Protecting public health, promoting patient safety, and reducing medication errors are

important priorities for FDA as an agency with the Department of Health and Human Services. As Mike Cohen mentioned, the Institute of Medicine in its December 1999 report "To Err Is Human" recommended that FDA shift the responsibility of testing proposed drug names to the pharmaceutical industry.

In November 2002, the HHS Advisory Committee on Regulatory Reform made a similar recommendation that FDA transfer in most cases drug naming safety testing to the drug industry, with FDA serving a role in reviewing data submitted by sponsors prior to approval of the drug.

The expectation would be that agreed upon methods would be used to screen for look alike and sound alike drug names already existing in the marketplace. FDA Commissioner Mark McClellan's emphasis on initiatives to improve patient safety recognized the important public health impact of reducing these errors.

Reducing the incidence of medication errors is not only an important FDA priority, but can

hopefully reduce the errors that hurt patients and contribute to the increased costs of health care.

Drug naming mix-ups, along with confusing packaging and labeling of drug products contribute to this ongoing important problem of medication errors.

It is difficult to put a firm number on how many medication errors result from named confusion due to under reporting of such events, but we know that a substantial number of medication errors are occurring because of look alike and sound alike name confusions.

Today we'll be discussing the current methods and approaches that are being used to screen proprietary names for similarities. We are excited about the opportunity to have not only an open public discussion, but to have expertise from the private sector, from the government, and from independent experts in academia.

We will be discussing issues related to methods of sampling, questionnaire construction, handwriting, and voice recognition models, the use of

expert committees, computer assisted analyses, how to conduct failure mode and effect analyses, and how some of these activities relate to efforts to do pre-market and to develop pre-market risk management programs.

We have a number of speakers who will participate in the open public hearing later this morning, and we have a public docket which is currently open, and we are expecting and accepting comments from many of you.

We hope you will take advantage of this opportunity to tell us what you think.

I believe it is safe to say that today will be the beginning of many more discussions in the future on this particular subject.

The questions posed on the FDA home page on May 30th request feedback on the methods that are currently in use. We are seeking information on what's currently being done in the private sector, what seems to work, and what doesn't and how to effectively evaluate and validate these current

methods.

The questions that were posed in the Federal Register are as follows.

Are current methods by sponsors and the FDA appropriate for evaluating look alike and sound alike names?

The second question: in studies to evaluate potential medication errors, what is the appropriate study design? What is the appropriate size of an expert committee? What is the appropriate size for a prescription drug study, whether it's for looking at written problems or voice recognition problems?

If you have an expert committee, what is the appropriate composition of such evaluators? How many physicians, pharmacists, nurses, consumers should be included? And what are the appropriate outcome measures to be used?

The third question focuses on what kind of information, such drug name, strength, quantity, directions should be included in verbal or

handwritten medication drug studies.

The fourth question refers to the issue related to risk management programs. Sometimes drugs are approved contingent on a risk management program. We wanted to hear examples of effective risk management programs that could be used to minimize look alike and sound alike confusion.

How should the effectiveness of such programs be evaluated?

And finally, should there be different trade name evaluation procedures for different classes of drugs, such as prescription and over-the-counter?

Once again, I want to thank our partners, ISMP and PhRMA, for their role in collaborating this morning in an open and constructive manner as we explore these issues.

With that I'm pleased to introduce our first panel of speakers this morning. Immediately to my right is Dr. John Jenkins, who is the Director of the Office of New Drugs and the center for Drug

Evaluation and Research. He will be bringing his office's perspective to the drug naming process.

Thank you, John, for being with us today.

Secondly we have two speakers from industry who will share the podium, Sharon Olmstead, who is the Executive Director and U.S. Regulatory Liaison for Pfizer Pharmaceuticals, and Tom Hassall, who is the Director of Regulatory Liaison for Merck. Ms. Olmstead and Mr. Hassall will be bringing the industry perspective to this particular issue.

And, again, thank you for joining us.

And finally, we will hear from Dr. Timothy Lesar. Dr. Lesar is the Director of Pharmacy in the Albany Medical Center, and he will be presenting the health care practitioner perspective, what he perceives to be the extent of the problem from someone on the front line.

One last sort of housekeeping note. The agenda today is very busy, and we're going to work as hard as we can to stay on time, and I'll be asking all of the moderators and speakers to keep close to

the agenda schedule. I think already we're doing quite well. I think we're about 15 minutes ahead.

Again, thank you all for coming this morning. I look forward to an interesting and engaging discussion this morning, and with that I'd like to turn the floor over to Dr. Jenkins.

DR. JENKINS: Thank you, Paul, and good morning to you all. It's really a pleasure to see so many people in the room this morning. This is the third talk I've given in the last two weeks in rooms about this size. The first talk there were 18 people there, and the second talk there were 15 people there. So my ego was really deflated.

(Laughter.)

DR. JENKINS: And so it's good to see a big audience again.

Paul had asked me to give a perspective on how we look at proprietary names as part of the new drug approval process in the Office of New Drugs.

And I see a typo on my first slide. We're not the officer of new drugs. We're the Office

of New Drugs.

So this will be a perspective for us in the Center for Drug Evaluation and Research. I know that the process is somewhat similar in the Center for biologics, but this is really the new drugs perspective from CDER.

Basically I can say that we considered the review of the proprietary name to be an important part of the review of any new application, and this review is performed by the New Drug Reviewing Divisions in my office, in the Office of New Drugs, and also in consultation with other offices within the center, including the Office of Drug Safety; their Division of Medication Error and Technical Support, DMETS, as it's sometimes called; the Office of Medical Policy. The Division of Drug Marketing and Advertising and Communication gets involved in helping us do these reviews, and we sometimes do these reviews with CBER colleagues as well.

The primary areas that we focus on in the new drug review process are really two. We look at

the name from a safety perspective, which is primarily to prevent medication errors, which is really the primary purpose of today's meeting, but we also look at the name from a promotional standpoint as well, to look for something we call fanciful names, and I'll try to explain that a little bit further by looking at what the regulations say.

But more importantly, we look for false and misleading claims that may be imbedded in the trade name or the proprietary name that we don't think are supported by the data.

Now, turning to the regulatory basis for our review, there's really only two citations that I'm aware of in our regulations. Maybe there are others. One is in 21 CFR 210.10(c), which says that the labeling of a drug may be misleading by reason, among other reasons, of, and number three of that list talks about the employment of a fanciful proprietary name for a drug or ingredient in such a manner as to imply that the drug or ingredient has some unique effectiveness or composition when, in

fact, the drug or ingredient is a common substance the limitations of which are readily recognized when the drug or ingredient is listed by its established name.

Now, this is kind of an odd regulation. It starts out reading kind of straightforward looking for fanciful names, but then it gets into when, in fact, the drug or ingredient is a common substance.

People who have been at the agency longer than I have tell me that this was originally intended to focus on things like maybe people wanted to cal their latest version of penicillin so you're really trying to hype your version of a commonly available drug to be uniquely effective or uniquely safe.

How this applies in situations where you have a new molecular entity is a new molecular entity a common substance? It's a little bit odd, but it does introduce the concept that we're looking for things that make the labeling misleading, and we do have the term "fanciful," although I don't think we use that as our basis in most cases.

The one that we rely on much more frequently, and this is the one that goes to the safety issue in that same part of the CFR, says that the designation of a drug or ingredient by a proprietary name, because of similarity in spelling or pronunciation, may be confused with the proprietary name or the established name of a different drug or ingredient.

So this is really the basis for our safety review for sound alike, look alike names, and it's important to note that it looks both at existing proprietary names and existing established names, and it really establishes a concept that first come, first serve.

So whoever is first on the market really kind of has the lead on those confusing names, and it's a principle we tend to apply. We compared new names against existing names. If your new name looks like it's going to cause a problem, you have to come up with a new name so that the existing name does not have to change.

Now, in the Office of New Drugs, when we look at proprietary names, obviously one of our major concerns is the issue of safety, and the focus of this review is to avoid medication errors, and we know that medication errors are a frequent cause of reports -- excuse me.

We know that confusion about labeling and drug names and packaging are a frequent cause of reports for medication errors. So we look not only at the proprietary name during our review, but we also look at the packaging and that includes the cartons, the container labels, et cetera, to see if they are easily confused with other products.

But we also look at the dosing instructions and how those instructions are written. For example, we look at the issue of whether there's a decimal point followed by a number and how that may be confusing as people start writing prescriptions for these drugs.

We do this review primarily in consultation with Jerry Phillips' group in the Office

of Drug Safety, and they help us to look for sound alike and/or look alike names to establish approved or marketed proprietary names and established names.

And I think Jerry is going to go over in much more detail later the methodology that his office uses to do that review.

Now, the other aspect we looked at, as I said, we look at the promotional aspects of the name, and the focus of this review is on that fanciful name issue that comes up in the regulations. But more importantly, we look for false and/or misleading claims imbedded in the name.

And some examples, superiority claims, suggestions that the drug in question is superior to other drugs for that same indication or sometimes imbedded in the name even though the data don't really support such a superiority claim.

Sometimes we see claims that are imbedded that suggest that the drug is effective for a different or an expanded set of indications than the ones that were actually going to approve in the

labeling. For example, sometimes we see names that try to incorporated suggestions about quality of life, that the drug is going to benefit quality of life even though the drug itself may be for a specific indication and has not shown a benefit on a quality of life measure that we find acceptable.

And sometimes the name may have imbedded in it claims for efficacy or safety that are not supported by the data, and sometimes we see that early on in your development you may have targeted an indication for the product, but when your studies come in, that indication really isn't supported by the data, and it gives us pause to consider whether we should approve that name, given that it has an implied claim that you don't have in your labeling.

We do this review much by ourselves, but we also involve our colleagues in the Division of Drug Marketing and Advertising to help us look for these promotional, imbedded, false, and misleading claims.

Now, there are other issues that we also

look at when we're looking at proprietary names, and one that's one of my pet peeves are the expanding proliferation of suffixes that get tacked onto drug names, and these are often confusing and subject to misinterpretation.

For example, we have the whole series of names that often are interpreted to mean something about some sort of a controlled release delivery system, such as SR, which many people would say is sustained release, CR, which some people say is controlled release, XL -- I'm not really sure how that fits into the release pattern, but it's often used on sustained release preparations -- XR and CD.

These are not well defined terms and can lead to confusion. We've actually seen some cases recently where a given product line may have multiple different sustained release or controlled release versions that may have different versions of these suffixes tacked onto the end of it that can cause confusion. Maybe one product is a 12-hour sustained release product and maybe one product is a 24-hour

sustained release, and maybe one is called SR and one is called XL.

You know, how are practitioners supposed to be able to keep those straight when they're making prescribing decisions?

We also see suffixes sometimes that may include implied claims. For example, going back to the XL suffix, does that mean that the drug is excellent? Does that in some way mean that it's better than other products?

And sometimes we see suffixes that may be misinterpreted as a dosing schedule, for example, QD or BID, and I've actually seen examples where the BID maybe was not a suffix, but it was actually incorporated into the name of the drug, and that was appropriate at the time that the drug was initially approved.

And then I've seen examples where the sponsor later wants to try to change that to once a day dosing. So you've got BID, which is commonly recognized as twice a day dosing, imbedded in the

name, and now the drug is indicated for once a day dosing, which can cause confusion.

The suffix area, we don't have a good standardized policy on this. So we see a proliferation of these suffixes across the various review divisions, and it's really hard to get a handle on these.

A couple of other areas where we focus. We look at the issue of multiple proprietary names for products with the same active ingredient, and by this I mean a given sponsor who wants to have the product indicated for different claims and decides that they would like to have a different name for each claim.

And we believe this has a potential for confusion to the practitioner as well as for the patient and can in some cases lead to overdosing, which can be a safety concern.

So we have generally discouraged use of two separate proprietary names for the same active ingredient for different claims, but we have

allowed this in certain circumstances, and the one that has been in the media most recently is Proscar and Propecia, where the Proscar is for the treatment of benign prostatic hypertrophy, and the Propecia is for the treatment of baldness.

Different dose schedules. Proscar is five milligrams; Propecia is one milligram.

So we do have a few examples where we've allowed this to occur, but we generally think it's a bad idea.

The other area that we look at is the same proprietary name for different active ingredients, and this is primarily something that comes into the realm of OTC products where you have a whole family of products that have a family name and then have multiple different active ingredients, and that's clearly a fertile area for confusion by consumers.

For example, the Robitussin brand name now includes multiple different active ingredients in those various products that are all under the

Robitussin brand name. Now, they usually have some sort of a suffix or additional name attached, but it is confusing.

You know, when I was growing up as a kid I always knew that Chlortrimeton was chlorpheniramine. If you go to the store now, there's probably multiple boxes that say Chlortrimeton, some of which don't even have chlorpheniramine in them. So it's definitely an area for potential confusion, but it's one that we have trouble getting a handle on.

Now, what's the review process that we follow in the Office of New Drugs? Well, first of all, we're willing to start looking at your trademarks and your proprietary names early. So we've indicated that we're willing to look at this as early as the end of Phase II meeting, and we definitely should discuss your proposed proprietary name at the pre-NDA meeting.

Something that's frustrating to you but is necessary because of the way the system works is

that our early agency feedback to you has to be preliminary in nature because there may be other products that get approved before you get approved that cause a sound alike/look alike problem that wasn't evident at the time we did our initial screening review.

So the old adage I like to remind people is your proprietary name is not approved until it's approved. So until you get the approval letter from the agency, we have not approved your proprietary name.

Now, we try to be reasonable in this regard. So we do try to do a final review of your proposed proprietary name by DMETS for sound alike/look alike within 90 days of the anticipated day of approval. There's always the possibility though that somebody could get approved the day before you do that's going to cause a problem.

Now, we try to identify those, and hopefully you try to identify those as well, but that's just kind of the nature of the first come,

first serve system that we're in.

And we also try to get DDMAC involved in reviewing the claim -- excuse me -- the proposed name near the time of approval because as I said earlier, this issue of false or misleading claims can be affected by what's the final label going to look like. So maybe all along it looked like you were going to get a claim for a specific indication, but then we decide that you don't have data to support it. So maybe that name then becomes problematic.

One thing to be aware of is that the final decision about the approval of your proprietary name rests with the Office of New Drugs. So the Office of Drug Safety and DDMAC are consultants to the Office of New Drugs, and we sometimes do consider their recommendations and decide not to follow the recommendations to reject the name.

So, for example, maybe we decide that the dosage form is so different or the settings of use are so different that even though there looks like a sound alike/look alike confusion potential, we think

the risks of that are minimal.

In other cases, we actually have reached agreements with sponsors where they agree to do educational campaigns to really get out the message about the difference between a new drug and an existing drug.

So we do consider the recommendations from our consultants very seriously. We sometimes disagree and our policy is that we should document in writing back to our consultants why we don't agree with their proposal that we reject your proposed name.

And in closing I'd like to give some suggestions that I can offer from the Office of New Drugs' perspective to you as sponsors. First, obviously, do your homework to avoid problems. And I know that you do this. I think Sharon is going to show a slide a little later that shows that you may even start talking about your proposed proprietary name before you even start your clinical studies for the drug.

That's obviously a good way to start, and as I understand it, you can monitor trademarks that are being approved by the Patent and Trademark Office to see if anything is coming down the pike that may be of concern to you.

The other recommendation is start early in your consultation with FDA. As I said, we're willing to entertain giving you a preliminary review of your name as early as the end of Phase II meeting, but keep in mind that that advice by the nature of the system has to be preliminary.

I would advise you to avoid imbedded implied claims, particularly those that are not supported by substantial evidence. We're going to pick those up, and we're not going to let them into your name most likely, and that's going to cause you problems at the end.

Don't put all of your eggs in one basket. So have several names available that you've tested and would have available, and we actually allowed -- and, Jerry, you can correct me if I'm wrong -- I

think we allow submission of up to two names that can go into the review for sound alike/look alike that Jerry's office does.

So consider submitting more than one name so that you have a back-up in case something happens with your preferred name.

And finally, it works best if you work cooperatively with us to try to resolve these issues when they come up. They do sometimes come up at the last minute. In some cases, that's the nature of the system, and it's best to work cooperatively with us to try to resolve those problems rather than getting upset and complaining and not being constructive.

We're looking to try to get your drug approved, and if we have a serious concern about a name, the best approach for you is to help work through that concern either by helping us understand why it's not a concern or how you can mitigate the concern in the marketplace through a risk management program or some sort of educational program or submit your back-up name so that we can move on and get your

drug approved.

Now, let me stop there and move on to the next speaker. Thank you.

(Applause.)

DR. SELIGMAN: Thank you, John.

Next I'd like to welcome to the podium our speakers from industry, starting with Sharon Olmstead from Pfizer.

Sharon.

MS. OLMSTEAD: Good morning, everyone. I have the task of trying to do my talk within about five minutes because Tom and I are going to be sharing the allotted time for the industry perspective, and those of you that know me, I can go on for much longer than five minutes. So I'm going to try to stick to my notes and keep it at five minutes.

I'm going to give a very brief overview of the industry perspective in terms of the entire trademark development process, and then I'm going to end my talk with just some of the regulatory

challenges we're experiencing or having experienced over the years that, in fact, this meeting today is trying to address. So I think we'll just bring those to the forefront.

And then my colleague, Tom Hassall is going to get up and talk about some of the DMETS reviews that have been posted on FDA's Web site as part of the approval packages that you can find when the products are approved. I think that will give some context to some of the comments I'm making today.

So I thought it would be interesting to put the drug development process and the trademark development process into a single slide so that you could get some perspective on how the two work together. So as you can see, on the top we've got the typical drug development process starting with discovery through to the launch of the product.

Now, imbedded in that process is the trademark development process, and as you can see, and it depends; each company is different. So it

depends on how early or how late the company feels they need to start the process.

But it can start anywhere as early as the end of the exploratory development phase up to, you know, the beginning of your full development, and then, of course, it ends with the product approval at which time hopefully you're successful and you actually have a proprietary name that you can go into the marketplace with at that time.

So as you look at the bottom list of names, in the blue section that actually represents the creation of the name. So I'll go through those steps briefly, and that's followed by the legal process. And actually Bob Lee is going to give much more detail to that legal process. I'll show you a slide on that, but he is our patent and trademark expert in our group. So I don't want to try and say that I know anything about that part of the process.

And then once the trademark has been established, then we go into the regulatory phase, which is the green blocks or arrows, and that's where

we typically do our dispensing testing and then go into the filing with FDA and look to get our approval with FDA with the name.

So the name has to come from somewhere. We don't typically pull them out of the air. So typically you have a strategic part that starts with looking at the brand attributes, and this may get to some of John's comments of why some of the names look the way they do, but you're going to look at your product and try to determine, you know, what the patient population is that you expect it to go into, what the disease state, how the drug will be administered.

And this actually contributes to the names that you're going to hopefully develop in this process.

We also take a look at our marketplace and where is this name going to be competing, who it's going to be competing with if there are other products in the marketplace at that time.

The next step includes a creative

component, and typically this is where the name is developed, and we rely on feedback from focus groups to help us in the development of our product names.

Now, moving from that process, we actually go into the legal process, which Bob will talk about, but those names that we've developed, these are what we would send forward to get legal clearance and get the trademarks established.

I think it's important to point out, as I had shown on the earlier slide, that this process takes quite a long time, and this legal process can take anywhere from 18 months to three years to complete and get your trademark registered.

I think it's also important as John had pointed out in the last arrow, the competitive monitoring, we do monitor what's going on, what our competitors are doing both with the trademark filings, as well as with the new drug approvals, and that's an important part of the regulatory process which I'll mention on the next slide. We want to make sure we're going to be able to maintain the name

that we start with and that we like.

I don't know if you can read that as well. It looked good on a small screen.

So the next step once we have cleared through the trademark process typically, we would go into the regulatory activities that would actually involve what we call the dispensing testing or error potential assessment or trademark safety assessment. I mean, there's a whole array of names that you can call this process, but this is where we would take our candidates that we've brought out of the PTO, or the Patent and Trademark Office, that have been registered, and we'll put them into the testing process.

And I know that there's going to be more discussion about the various steps in the process. So I'm not going to get into that because I want to talk more about the regulatory challenges that once we have these names that we like, how to go about getting those names approved.

So we generally can go into the testing

component with anywhere from 12 to 15 names, and in an ideal situation, and the recommendations that come out of that process hopefully we'll have a pool of three to five names that we can choose from, and generally one of the bigger challenges that we find is our senior management and the names that they fall in love with that we maybe can't use based on the data that we've gotten back from the dispensing testing.

So from the regulatory and legal process, that's the big challenge that we find. So once we have that pool of three to five names -- and one thing I will add is that as I'm walking through this process, this is one way of doing it. I know different companies do it differently and some may agree with this process and some may not. And I'll try to point out some of those differences.

Once we have our pool of three to five names, typically you will then go ahead and file one to two of those names with FDA for your preliminary approval.

Now, depending on the company, some have the philosophy that they prefer to only submit one name because they feel it's their choice and they want to go forward with the name that they've worked hard to bring forward.

Other companies feel that they would rather not put all those eggs in one basket and have a fall-back name. And I've actually worked with both scenarios, and I can tell you that when you do go forward with two names, you can be assured that just because you have two names it does not mean your first name will be automatically rejected. It actually does get tested, and it is considered. And the second name just becomes the fall-back if in the event that it is not accepted, then they would go on to look at that name. And it can be helpful to have that second name up front, and I can share a little bit of an experience as we get further down.

So going through the process with FDA, typically you can submit that as early as the end of Phase II, and that process can take from several

months up to a year depending because, of course, the names that are in the queue for the NDA reviews clearly take priority over those that are being submitted for the early review.

And once you have that preliminary approval, it gives you some sense of where you name fits in with the currently approved drugs. As John said, it's not a guarantee. It's not a final approval, and of course, you could run into some difficulty once your NDA is submitted if other names have come through the process once you've gotten your preliminary name.

So as you move forward into the final steps of your product development and you're ready to submit your NDA or your BLA, you typically have -- hopefully you have your preliminary approval in hand and you provide whatever necessary information you feel would continue to justify your name in with your NDA submission.

However, there are occasions where you may want to test your name again looking at what has

changed in the environment since you received your preliminary approval to make sure that your name is still hopefully safe and you should end up with it at the end of the day.

In other case sponsors may want to do additional testing because they may still feel really tied to that original name, but they received a reject in the preliminary approval. So they may want to try and supplement and try and change FDA's mind with additional, more intense study of the name and try to turn it around.

And then finally as you reach the approval stage of your NDA, a couple of things can happen. Your name can obviously be approved and accepted and yo move forward.

Your name, there may be some questions about your name, and recently FDA has begun implementing risk management plans, which we actually find very useful because then it does not result in the what we would call the worst case scenario where you have to go and find a new name at the end of the

day. I'll talk a little bit more about that at the end.

So having kind of walked through the process in a very high level overview, here's some of the regulatory challenges from an industry perspective that we see, and the first one is the predictability of the current model that not only FDA is using but also the commercial vendors that many f us contract with now.

And currently as we understand it, the methods that are being used have yet to be validated. This raises the question are we actually testing what we think we're testing. So are we actually measuring the potential of the air or not. So I think that's the first step that we need to be considering.

And then secondly, if we can build in predictability and validation into our model, it will help us to understand what level of evidence is necessary to achieve FDA clearance because right now when we go forward, we have our package of

information when we submit it, but we don't have 100 percent assurance. I'm not even sure I could say that we have 75 percent assurance that we're going to actually achieve that name at the end of the day.

So I think this component would actually help, and I think that some of the questions that we're answering today will actually get at the heart of this issue.

The second item I raised is the error threshold, and I think in the past, and I think this is changing, but the perception from industry has been there has been a real zero tolerance when it comes to name similarities. So that if your name has some hint of similarity with another name, regardless of the public health impact, that name would actually be rejected.

And so for us that's difficult because it did not at the time give us an opportunity to try and address whether or not that similarity would actually result in some public health problem. So I think going back to FDA's recent acceptance of certain risk

management plans and educational programs, it allows us to do more of a risk benefit of the name and those potential similarities where we might actually be able to put a name out into the public domain and monitor it and manage it and hopefully not have any problems.

But, again, I know that in some cases sponsors have agreed that after there's actually a threshold that's established of medication errors, that it's a certain number reached, then the name would have to be changed.

And then the final -- my time is up -- so the final thing I would say as I get ready to introduce Tom is the ever ending or never ending train wreck scenario which is where you get to the end of the day and your name is rejected, and I think many of the issues that we're going to talk about today will help to address that and move us forward into finding alternatives to last minute name changes.

Thank you.

And now I guess we'll introduce Tom Hassall.

(Applause.)

MR. HASSALL: Good morning. First I want to thank FDA, PhRMA and ISMP for putting on this workshop and particularly for granting me the opportunity to speak today.

Because we have a lot on the agenda today and I only have a few minutes, I'm going to try to keep to the schedule also and give us time for some useful discussion.

Second, I want to emphasize that the issue of how to effectively evaluate trade names to prevent medication errors is not an easy issue. Sharon has outlined the extensive effort that most companies today put into the selection of trademark, yet both the EMEA and the FDA still find about one third of all the names they review to be unacceptable.

It's hard for me to understand how well intentioned people with a common goal of a unique

mark can be this far apart in their conclusions, and clearly we have to find ways to narrow this gap. I think this workshop today is a great start for doing that.

To do this I think we need better methods that yield, number one, reproducible results, and we need some standards against which to evaluate the results in order to improve the predictability of the outcome.

To set the stage for this, what I'd like to do is summarize the survey that I did of 22 FDA trademark reviews, and I really did this because as I got more and more involved in this issue, I felt like I had a need to understand what was going on in terms of FDA's review, how it was being conducted, what it consisted of, and to get some sense of how outcome seem to pertain to the methods that were used.

So first let me start by taking a look at the FDA process. FDA's trademark evaluation process involves three steps. There's a panel composed of the DMETS staff who review the name against the

number of standard compendia and references, as well as their own experience, and come up with a pool of existing proprietary names, and of course, nonproprietary name that maybe be confused with the proposed trademark.

And then there's the prescription analysis studies, which according to the text of several DMETS reviews is intended to, quote, determine the degree of the proposed name with other names due to handwriting or verbal pronunciation.

And finally, there's the safety evaluator risk assessment in which a reviewer considers the pool of names identified by the expert panel, the results of the prescription analysis study, and potential mitigating factors, such as intended population, dose, dosage form, regimen, route, consequence of the error, and others.

And this judgment call is subjective and leads to the conclusion of the review. It's subjective, although clearly based on the earlier parts of the review.

I obtained 22 reviews just off of the FDA FOI Web Site and looked at those to get a better idea of the final review conclusion and how it related to the outcome of the prescription analysis studies that FDA conducts.

In my sample, I had 22 reviews, and I should say, by the way, I don't consider this to be a scientific survey. I think Jerry's group does something like 300 reviews a year. I looked at 22. So, I mean, it wasn't intended to be some kind of a scientific survey. It was really just to give me some idea of how the process works.

In five of the 22, there was no prescription analysis. So I concentrated on the remaining 17. You can see the sort of spread. Most of them were from 2001 and 2000. There was one early one and a couple of those from 2002.

The reviews involve nine different reviewers, and the average time to completion was 60 days. So I think we can give Jerry's group a hand. They've generally said that they get these reviews

done in a couple of months, and by golly, at least from my little survey, they do.

As you can see from this slide, a prescription analysis study consists of three parts" the written out-patient sample, the written in-patient sample, and a telephone order which is left on a subject to voice mail. They do their survey within the FDA's staff.

And approximately 30 people are included in the sample size of each of these phased, 30 out-patients and 30 in-patients, 30 verbal.

The response rate in the survey that I did was about 60 percent for the written orders, somewhere around 18 or 19 people responded, and about a little under 50 percent for the verbal orders.

Now, the third column shows the percent of responders who correctly identified the proposed trademark, and I find this to be a somewhat meaningless statistic actually when I got looking at the reviews because actually the vast majority of these incorrect responses are phonetic variations of

the actual spelling of the trademark and, therefore, may be of very little consequence.

What's more important, I think, are the misidentifications with an existing product name, and there are four of these. Of the four, FDA found two of the trademarks acceptable in spite of the mix-ups while it concluded that two were acceptable.

Let me try and explain this slide a little bit. This compares the final review conclusions with the outcomes of the prescription analysis surveys conducted, and what I did is I tried to get some way at getting at sort of a total score of correct versus a total score of incorrect, and just very simply essentially added up the percentages of correct responses in a survey versus incorrect responses to see how the balance came out.

And so reading horizontally you can see that seven names were deemed unacceptable out of the 17. In five of these seven prescription studies, the incorrect answers exceeded the correct answers. So that sort of on balance is the way you would expect

it to go.

Similarly, again, reading horizontally, in ten instances where the name was judged to be acceptable, there were more correct answers than incorrect, although the spread is not as wide, and obviously if one had fallen the other way it would have been a 50-50 split.

Reading vertically on the incorrect greater than correct column, of those nine tests with more incorrect answers than the correct ones, almost as many trademarks passed as failed.

While I recognize you have to be careful about drawing conclusions from such a small look at these reviews, actually there's good news here and bad news. I mean, I think the bad news from the industry standpoint is that it sort of confirms our sense of unpredictability.

The good news is it's clear that FDA's safety evaluators, when they do the third part of the review, are not just taking blindly the result of the prescription analysis studies and are, in fact,

considering the other factors or mitigating factors that might allow a name to go forward.

So now to my conclusions. I've done a lot of thinking about this and particularly the small numbers in the surveys going in and also the small number of responses. I've thought a lot about what I consider the irrelevance of incorrect responses that are merely phonetic variations of the proposed name and also about the fact that we don't really seem to know prospectively what we're looking for in the studies themselves to declare a win or a loss.

So my conclusions are that prescription analysis studies don't really test the name for the risk of medication error. I don't think that an incorrect response involving an existing name is not significant or is significant by itself.

I don't think that a lack of an incorrect response involving an existing name is significant by itself. I think the prescription analysis studies do not determine the degree of confusion of the proposed name with other names due to handwriting and verbal

pronunciation as stated in many of the reviews, and I don't think these tests necessarily produce reliably reproducible results because similar tests that sponsors have contracted for prior to the submission may come up with different conclusions, and in fact, they do.

I think prescription analysis studies are useful as screening tools, and I think what they do is they enrich the pool of potentially confusing candidate names that's initially generated by the expert panel and that will undergo the safety evaluator risk assessment.

So, on the one hand, you have a small body of the expert panel who comes up with names from looking at compendia, and in a sense this is just really a bigger expert panel that draws on a wider experience.

I also think that prescription analysis studies do not identify potential errors with a higher or lower risk of occurrence than the other names that have been put in that pool from the expert

panel. So, in other words, the potential look alike and sound alike names identified in the prescription analysis studies should not be given more weight in the safety evaluator's risk assessment than the names in the pool identified by the expert panel.

Recommendation 238 of the HHS Advisory Committee on Regulatory Reform called for FDA review of manufacturer generated data from protocols designed to evaluate their products' names for a possible look alike and sound alike names.

To avoid a problem with this recommendation, I think it must be realized that we have not as yet identified any method for reliably testing trademarks. We must avoid interpreting this to be a recommendation or this recommendation to be a call for a specific test as opposed to a recommendation for a predefined plan or protocol that the company intends to use to evaluate the proposed mark.

Interpreted in this way, I think it should then be FDA's role to agree upon what I have

sort of called good naming practice, and since those would be called GNPs, that has a very great look alike satellite to GNPs.

(Laughter.)

MR. HASSALL: But their role should be to essentially define good naming practices and then assess whether or not those practices have been followed in the selection of a trade name and thereby building quality into the process.

And my final conclusion is the usefulness of any test or study that's purported to actually assess the risk of name confusion that may contribute to medication errors must be validated before it can be recommended for regulatory purposes.

Thanks.

(Applause.)

DR. SELIGMAN: Thank you very much, Sharon and Tom.

Finally, our last speaker on this panel is Dr. Timothy Lesar, who is the Director of Pharmacy at the Albany Medical Center.

Dr. Lesar.

DR. LESAR: Good morning. I'd like to thank ISMP, the FDA, and PhRMA for inviting me to speak to you today about the practitioner's perspective on the problem of the look alike/sound alike problem with medications. And I will come from this with a little bit different perspective obviously, but I'd like to mirror many of the things that were said about risk assessment and determination of problems.

I'd like to go through this by actually giving some idea about the evidence base for some of my comments are, and then talk about really a conceptual framework from a practitioner's standpoint about this issue, give you some select examples, real life examples of our problem that we see very commonly every day in our institution, and then some of the implications that I have for risk assessment and safety enhancement.

At Albany Medical Center, we have systematically collected medication errors since

1987, evaluated them and determined causes, contributors, tried to develop a framework for understanding how these things occurred and what increases risk or decreases risk in order to improve our patient safety.

This includes over 30,000 prescribing errors alone. As I said, we look for contributors, confounding variables, and what appears to be the underlying cause or one of the contributors, and out of these errors, about one in five is related to nomenclature issues.

This is just one example. This is from an article we published in Journal of General Internal Medicine last year, which demonstrates just related to dosage form naming and nomenclature over a twofold increase in five years, from 1996 to 2000 related to dosage form and nomenclature. About 70 percent of these errors are specifically due to the name of the drug and suffixes as Dr. Jenkins had mentioned.

So we have a large database, which we can

understand these errors. What we have found is that from our perspective is that drug names clearly have a clear potential for error. From a practitioner's perspective, we can often look at a name and go, "Oh, my gosh, that's going to be problem here or there."

We can see them commonly cause or contribute to patient harm. At our institution that we know of, that we know of is all; we know that there are more, but there probably are many more. At least two errors a day occur because of nomenclature problems.

And I will say that there is a perception often when we see some drug names come out or in suffixes used, there is a perception that that safety is not the primary consideration in product naming.

The other thing that I wanted to stress is that often from our perspective very simple changes can make dramatic improvements in safety. So very minor modifications in doses, names, those type of things can improve safety markedly over what might be an approved name.

Trying to develop a conceptual framework for sound alike/look alike problems is not easy because as you have already heard, this is a very complex problem. Our concept is that you take a product with a name, which should tell you everything about a product. You insert it into our medical care system. It interacts in this complex care system and outcome problems and errors, many of which are predictable. Some of them it's surprising until you think about it, and you say, "Well, that was predictable, what was going to happen."

And then how these errors and interactions occur will depend on the specific product characteristics and all of its characteristics, as well as all of the characteristics of the care system involved. So you might see slightly different things in an in-patient setting than you might see in a community pharmacy setting.

But typically in hindsight, errors occur in quite predictable ways, and since we've had all of

these years of looking at errors, we can often look at a name and say this is the error that's going to happen with this particular product, and indeed, not too long after we see them.

Obviously there's an issue related to the risk for error versus the risk for an adverse drug event. As many people know, errors happen all the time, but they don't very frequently produce ADEs, that is, a high percentage of them will not produce an adverse drug event for the patient.

But a lot of things will be determined. The risk for actual adverse drug event, you know, what the error specifically is, what the drug, those type of things, so there are different things that we look at in terms of determining what the actual patient risk is rather than just the risk for error. Sometimes you're surprised at how risky it is even things that you thought were not going to put patients ar risk.

And so from this conceptual framework the way we would look at this is that any or all

characteristics of a drug product can increase or decrease the risk for error. It must be considered in risk assessment.

And so all of those factors, generic name, brand name, dose strength, frequency, where its used, who it's used in, all of those things are very important in determining what the potential risks are.

And so our conceptual framework is this. You have a product that you insert into this tornado of things going on in our hospital or any medical care system, and eventually out spins an error. And I tried to show this. It's a little bit difficult to read, but I just started listing all of the different types of things that might be related to problems in determining nomenclature.

And so, again, we're inserting products into this vortex of things going on in health care, other products, processes, knowledge deficits, all of those types of things that are problems and processes within our organization, and eventually we're going

to see errors.

Again, these things are all mixing up in different ways and different times. However, many times you can pretty much be sure that you can understand how errors occur.

I'd like to show you some examples. this is a case in which Humalog -- about any practitioner that asks, the name Humalog can cause confusion, and I would tell you that most practitioners would say yes, and in this case, Humulin Log (phonetic) was actually ordered as Humulin Log. So you have a rapidly acting insulin. The physician thought Humulin L was the same as Humulalog, writes for Humulin Log, combines the two names, 85 units, a very high dose for a rapid acting insulin.

The nurse, of course, thought they meant Humalog. So you can see this was just a little term. We see these types of errors all the time with this class of drugs. Very predictable.

To spread the issue on to the competitor, NovoLog, the other rapidly acting insulin, and in

this case we got a call from a nurse who asks, "NovoLog, that's regular insulin, isn't it?" Because, you know, when you have insulin R or Novolin R, that means regular insulin.

Well, she misinterpreted the registration R in a circle to mean R, to mean regular. So you can see you combine knowledge deficits with someone reading a page who doesn't know. This is what happens when you insert these things into that vortex of medical care.

Other examples. Another predictable. It was not unpredictable at all that OxyContin was going to be confused with its generic name oxycodone. In the top example you see where the physician wrote oxycodone when he should have written OxyContin in a very high dose, and when he wrote OxyContin 5 milligrams, that was supposed to have been the oxycodone.

That was not too much of a stretch, but if you notice the example below in which the physician orders OxyContin, 60 milligrams when

actually he meant MS Contin. Again, serious overdose of narcotic analgesic.

Dr. Jenkins already mentioned suffix is something we've complained about for many years, and these are two new products. I believe actually I made that up. Cardizem is actually LA, I believe, but that's the third sustained release Cardizem formula.

I'll tell you it's just a matter of time. We haven't seen these products in our institution, but I will tell you as soon as we see them, we will see errors.

Add in the component of legibility, I know that's going to be discussed today, but top example, vancomycin becomes Unasyn. Protamine becomes Protonix. Now, that name confusion wouldn't have occurred when we only had oral Protonix, but now there is injectable Protonix, and indeed, a typical case. Patients are seen in similar environments in which Protamine -- and this was actually given in this case -- Protonix was given instead of Protamine.

In the bottom, Capoten for Cozaar. So what you can see is that we commonly -- this is just a few examples of the kind of things every day that practitioners are faced with.

New to us over the last number of years is the introduction of newer technologies. This is a screen from our Pixus Unibase cabinet in which a case in which a nurse went to retrieve Lopressor, was using the brand name screen, was going to obtain Lopressor, overrode the Pixus machine controls, misread, pushed Levophed, out drops the bowl with just one vial in it, and indeed, those vials as you can see in the lower right-hand look very similar. Took the Levophed and gave that IV push.

So we have this interface, and this isn't the only type of interface with technology that now creates a new complexity for the problem of look alike/sound alikes.

And then one final point is how the importance of the entire drug product is important. In the upper example, the anti-hyperlipidemic drug

Tricor was written as Tricor 125 milligrams. The pharmacist, trying to figure out why this was, trying to tell them that that's not the correct dose for Tricor, finally sorted it out that indeed it was Tracleer, the drug for pulmonary hypertension. And so that was caught because the doses differed significantly

In the example on the bottom, it was where you see the two lower Proscar, which is a drug used for prostatic hypertrophy, and then Prempro. Well, most people on Prempro don't have prostates. So --

(Laughter.)

DR. LESAR: So it's interesting that those are exactly the same. The pharmacist initially just missed this completely. It was so easy to do. It was supposed to have been Prinivil.

So you can see that from our concept the entire drug product, the interface is extremely important. So the implications to us as practitioners are that errors are generally

predictable. They will surprise us, but they are typically predictable, and they can be used successfully in error reduction.

This predictability can also be used to enhance safety. So when you have a Cardizem LA, make it 245 milligrams instead of 240 milligrams. Clinically insignificant, but at least it's a red flag to the pharmacist to say, "Hey, they meant to use a different dosage form."

So enhancing safety is not that difficult. When one considers all product characteristics and also what is the environment in which it will occur.

So to summarize, drug names, labels, and packaging are a major contributor to medication errors, a problem we see every single day, makes our work much more difficult.

The risk for errors is determined both by the product as well as the environment it will be used in, and that risk assessment must include multiple drug characteristics as well as what care

systems they will be used in.

And, again, to reiterate the point that small changes, understanding these processes allows us to make small changes which will enhance safety.

Thank you very much.

(Applause.)

DR. SELIGMAN: Should we move on to the next panel?

Thank you very much to all of our speakers this morning for setting the stage for what I hope will be a very lively and challenging discussion today. Thank you.

(Pause in proceedings.)

MR. LEE: Well, it has been said that medicine is a blend of art and science, and how much art and how much science is always an ongoing discussion, always an ongoing debate.

I think trademark development, the legal searching, and the safety evaluation of the trademark manager also are a blend of art and science. We're all about to learn more as the day goes on about the

mix of art and science as we hear from a group of panelists who are deeply involved in the process.

It's my pleasure to welcome today the panel that we've arranged:

Jim Dettore, President of Brand Institute;

Clement Galluccio, with Interbrand Wood, and, Clement, I think you're in the RxMark group.

MR. GALLUCCIO: That's correct.

MR. LEE: Susan Proulx, President of Med-ERRS;

Jerry Phillips, FDA, DMETS;

And Toni Stifano from CBER.

This is just an overview of what we'll try to accomplish in this panel, really looking at the current methods that are used to try to assess medication error potential.

Just two slides that will show what many companies do in parallel when they're doing a name selection, name clearance. You've heard a lot about the trademark or some about the trademark legal

clearance versus what's done in order to try to evaluate trademarks from a medication error perspective is trademark legal clearance, and by legal clearance, it's really trying to predict ahead of time what the test mark's likelihood of confusion will be with other trademarks that are already on the marketplace.

This is something that all trademark people do for over a century. So the concept of being concerned about look alike and sound alike similarity in trademarks is not a new one. Names are not just selected out of the air and you hope that when you get into the marketplace there won't be any confusion.

From a legal point of view, you have to be careful. If it's likely to cause confusion, there are legal rights that other trademark owners have, and so you have to, from a legal point of view, avoid infringement, and that's done by looking at the similarity of the marks that you're testing, and we do that by looking at the USPTO trademark database.

We look at common law references. We search for domain names at the same time because we may want our trademark to be part of a domain name, and we search other sources that are available to us to try to make sure that the name that we're picking is not a name that is already being used in the marketplace and that we're likely to cause confusion with.

As you do this, you get a frame of reference about the level of similarity that you must tolerate with all of the product names that are in the marketplace in any industry, and that's also true in the pharmaceutical industry. Because we often see that what we want to do is look at that similarity that may cause medication errors. That's much easier to say than to do.

You have to have a pragmatic sense about how much similarity is acceptable and how much is excessive, and it's very difficult to predict.

I was struck by the example of Capoten and Cozaar that the last speaker mentioned. Every

mix-up is identified by two product names. That's how we know how to identify the mix-up.

That does not mean that the product names caused mix-up. Something else could have caused the mix-up, and we identify it by mentioning two names.

From a trademark attorney's perspective, Cozaar and Capoten just are not confusingly similar. It's hard to say you could predict that mix-up.

Now, if you look at handwriting and the handwriting is very, very bad, you could almost mix any two things up, I suppose.

Promising trademark candidates that survive the legal search, so after going through that process -- let me go back just a minute. After yo go through the legal screen, you come up with a subset of marks from the universe of all marks. You come up with a subset of marks that have more similarity than the marks that are left out of the subset. That doesn't mean they're going to cause confusion.

So you have to do a legal analysis. You look at those trademarks and you decide are those

likely to cause confusion in the marketplace, and there's no objective test that one can use to do that. It's the sense; it's the feel of years of practice. I guess that's why they call it the practice of law. You have to practice it.

And we get that opinion from a trademark attorney, and then we have these surviving candidates with what is an acceptable level, in our view, of similarity because you can't eliminate all similarity.

By way of example, we have more than 26 products on the marketplace. That means some of them are starting with the same letter. So there's a level of similarity, and you might say, well, that's acceptable. Well, how about the first two letters or the first three letters, and so on and so forth; where is that level of similarity?

Then you go to once we take these marks that clear, we go through another process. I say it's done in parallel, but first you have to get the legal clearance, and then you go through and you say:

how do we test for a medication error analysis?

And right now most of the methods are look alike/sound alike prescription testing. There's responses to questionnaire, and you saw reference to the way prescription testing is done through handwriting and verbal orders, handwriting tests and verbal order tests.

Then response to questionnaires. We consider medical terms and abbreviations, other dispensing issues, clinical setting, dosage. More and more we've been doing that in response to medication errors.

Then we get a subset of names with error potential, names that are identified as possibly causing, could cause errors, and we ask, many people ask experts, that is, pharmacists and others what do you think about the possibility of this mark coexisting in the marketplace from a safety perspective. And the ones that survive that analysis are the marks that go forward for review by the FDA.

Now, we've broken up this session into a

number of different areas, and these are questions I'd like to ask the various panelists here who have different ways of testing for error potential. Some of these questions, for example, Jim, if I could ask you the first question.

How do you select your respondent sample when you go out to get data on error potential?

MR. DETTORE: Yes. Thank you. Thank you, Bob.

We randomly sample through a prerecruited panel of practicing pharmacists. We make sure in the process it's designed to achieve the representative sample based upon the product information itself and the prescribing profile, as well as the geodemographic characteristics in order to minimize sampling error.

The design itself is actually administered by our staff internally of professionals who have experience in the areas of survey design.

So it's randomly sampled.

MR. LEE: Jerry?

CAPT. PHILLIPS: Well, my sample size is of the FDA physicians, pharmacists, and nurses, and we basically ask for volunteers within the FDA to volunteer to do these studies. So the composition, it's not random. It's strictly based upon what we have internally to work with. So they're a group of physicians, pharmacists, and nurses from all components of the FDA that participate in the studies.

Obviously we cannot go outside FDA because these are confidential applications.

MR. LEE: Well, when you go out to get the data that we often talk about with prescription testing, you are sampling. You're going out and asking practitioners for information about how the handwriting might cause errors or how the verbal orders might cause errors.

So who do you include in the same? And I'm going to direct Mr. Clement. Who is included in the sample of people that you ask?

MR. GALLUCCIO: Well, let me begin by

saying for each assessment we have two distinct sample components. The first sample component is a quantitative primary research study that represents the profile of health care professionals who we anticipate will either prescribe, dispense or administer the product.

For the vast majority of projects, we include physicians specific to the profile of the anticipated prescribers, nurses specific to the anticipated profile of the dispensing environment, and pharmacists and other dispensers, such as unit clerks that represent a cross-section of dispensing environments, primarily hospitals and retail.

For example, a hospital only product would also be validated by retail pharmacists to reflect the larger number of products as opposed to the products only prescribed within that environment.

The second sample component is a qualitative primary research study managed by an independent consultant. We've worked with Dr. Neil M. Davis, of Safe Medication Practices Consultants,

for quite some time, and his studies conducted with individuals who share his interests in minimizing medication errors. So naturally a bias is introduced, but the benefit is that you are interfacing with individuals that are attuned to the nuances of medication error as was alluded to earlier.

MR. LEE: Thank you.

Sue?

DR. PROULX: We have a database, Med-ERRS has a database of practitioners that include pharmacists, nurses, physicians, other health care practitioners, as well, but based on the process that we use, we primarily use pharmacists because of their greater knowledge of medications.

Depending on the product, if the product is going to be used in a hospital, we tend to use hospital pharmacists. If it's an oncology product, we will try to enlist the aid of oncology pharmacists because, again, of their greater knowledge.

MR. LEE: What size sample size is often

used in your group?

Jerry, I'm going to direct that to you if I may.

CAPT. PHILLIPS: Well, we have about 130 folks that interpret the prescriptions, and as you saw from Tom Hassall's presentation, those are divided. The 130 are divided, one third into the written prescription studies, one third into the out-patient, and one third into the verbal orders. So there are approximately 30 or 40 potential candidates that review those trademarks at a time for each particular portion of the study.

MR. LEE: Jim?

MR. DETTORE: This is probably one of the hottest topics from at least the sponsor's standpoint to clients. I know they're continuously asking probably Interbrand Wood, Clement, and Susan and other vendors who are here today. I know it's one of a top topics for us.

We typically recommend a minimum sample size of 200 to be disbursed equally between 100

pharmacists and 100 physicians. We feel this achieves, one, representative sample of both the physicians and pharmacists. Out of the pharmacists, we usually go 50 percent in hospital, 50 percent retail to make sure that we try to identify as many, as they say, cast a net as wide as possible in order to try to find every possible source for medication error potential for sound alike/look alikes.

MR. LEE: Well, those questions were to try to give you a little insight as to the particular numbers and features of the prescription testing process, handwriting testing, verbal order testing. That gives an idea of how the data is collected from what we might generally call prescription testing.

There's another aspect to gathering information about the new test drugs besides the prescription testing, and that is certain groups will also ask questions of the respondents to gather additional information beyond the testing, and so we wanted to direct a few questions to the panelists about the manner in which questions are asked.

So I'd like to look at the first guide there, and those who do use questions in the process, are questionnaires self-administered by the respondents?

I'm going to direct that to Sue.

DR. PROULX: Yes, they are self-administered. We put together a short survey on our Web site over the Internet that practitioners complete. Every so often we'll have to do a specialty type questionnaire because we're looking at a particular problem for a client, and then we may have that done by E-mail with our practitioners.

What we do is we have the practitioners, when we send them a message asking them if they're interested in participating in a particular project, and if they agree to participate, we take a sample of them to actually be able to log into the particular survey.

We give them a time limit, and then after that date and time, the project is shut off.

MR. LEE: Clement.

MR. GALLUCCIO: Well, what I would share is that an important consideration for many of our clients is the fact that they're seeking to develop global brands, and when you consider that you have to select methodologies that you can apply consistently from market to market.

To recap per my earlier statement, each assessment contains these two components, primarily qualitative and a primary quantitative. To begin with the quantitative, you know, we use primarily a mix of telephone, voice mail, and fax methodology. The interview is conducted by a professional market research interviewer.

However, from time to time we will also integrate a face-to-face interview or perhaps an on-line survey, and this is specific to prescribers as well as nurses.

For dispensers, as well as other individuals, such as unit clerks, and so on and so forth, we use a secure on-line self-administered survey. The methodology is perfect for sharing

visual as well as verbal stimulus.

For the qualitative component, basically what we use is a secure E-mail communication, and that is collected and synthesized by SMPC.

MR. LEE: Thanks.

Jim, let me direct the next question to you. Do you use personal interviews in your process?

MR. DETTORE: Yes, we do, Bob. We use an external professional review committee made up of health care professionals from around, again, the U.S. They are assessing both personal interviews as well as round table discussions to discuss the issues of medication errors within the research area that we conduct for our clients, and this provides a qualitative supplement to the quantitative analysis for interpretation studies that we conduct for all projects.

MR. LEE: Sue or anybody else, personal interviews?

DR. PROULX: Yes. We don't generally use personal interviews on a regular basis, but if we

have questions to our practitioners who have answered our surveys, we can call them or E-mail them for follow-up information.

And in addition, we will contact oftentimes specialists in that area who are not necessarily participants in that survey but someone who has a knowledge of that particular area where that product may be used, and we try to get additional information about how that product would be used in its drug setting.

For example, if we were working on a product for a radiopharmaceutical, we would contact nuclear pharmacists that we have that we're familiar with, that we know, and ask them questions about how a radioactive nuclear product would be used. It's something that's not necessarily our area of specialty.

CAPT. PHILLIPS: And I guess it's important to realize that FDA does not use questionnaires. It's not part of our assessment. So that's why I'm being silent.

(Laughter.)

MR. LEE: Thanks, Jerry.

Clement, questions open ended or multiple choice?

MR. GALLUCCIO: Well, we use a combination, and I think that specific to the identification of the candidate when written or when spoken, it is an open end response, and the one nuance that I would add to that dynamic, I believe the trend for the most part has been over the years to use a singular set of stimulus. So, for example, you will see a cross-section of the candidate written in five different expressions, and the respondent will attempt to interpret what is being communicated given those five different expressions.

And I think that what we have learned over time, since certainly that is where we began, is that there is certainly much value in using multiple sets of stimulus. For example, if you sample 200 people perhaps sharing anywhere from 20 or 30 different sets of stimulus with those 200

respondents, so this way you more accurately replicate what is being communicated as opposed to the bias of that one single set.

However, we do ask other questions related to the candidate. For example, an overall assessment would be a multiple choice. The ability of the candidate to be communicated clearly in the context of that specific environment, we will use a Likert scale.

So we use a mix of different types of measures and methodologies.

MR. LEE: Anybody else want to respond to open ended versus multiple choice? Jim or Sue?

MR. DETTORE: We use both. We use open ended and also multiple choice at the same time. I think the direct multiple choice and at the same time feedback by open ended gives a broader understanding of potential for medication or at least input on sound alike/look alikes for information.

DR. PROULX: We use open ended question only because we want to get as much information as we

can from our practitioners, and I think the types of questions we ask may be a little different from my colleagues. We're asking them what they think this particular handwriting or verbal sample could sound or look like versus what do you think this says.

So it's a little bit different, and because we want to elicit as much information as possible, we get a lot of information back from our practitioners based on open ended questions.

MR. GALLUCCIO: To that point, if I just may add, one of the interesting bits of learning that we have collected over the years is the fact that we do need to be careful in not leading the respondent to identifying other marks that may be perceptually similar; however, in that clear identification of what is being communicated, marks that would not be identified.

So to manage that, I guess we approach it in somewhat of a different way. We actually have the identification of the stimulus occur unaided. So it is simply the proposed trademark. And once we share

the product description later in the survey, it's at that point we begin to ask a more global, inclusive question: what is your overall assessment and what is your rationale for your assessment?

And if there is a strong perceptual similarity, that will be communicated within the open end response.

CAPT. PHILLIPS: I was thinking as we look at responses back from the prescription studies of the 130 people from FDA, there will be unsolicited comments at times from some of the participants about similarities that exist. It's their perception.

So although we're not structuring our questionnaire, at times we do get that feedback. That's different than a strict interpretation of the prescription.

MR. LEE: Sue, I'm going to direct to you. Do you supplement the respondent input with other data? Do you do computer searches? What databases might you use? Do you look at clinical information?

DR. PROULX: Oh, absolutely. We really believe that the practitioner responses is just the beginning of the process, so to speak. The big thing that we do is perform with our expert panel a failure mode and effects analysis, or I guess we're calling it MEPA today. It's the new abbreviation that I just learned this morning. I guess we all did.

Is that a failure mode, Jerry? MEPA is now FEMA or FEMA is now MEPA, or whatever.

And I know we're going to have an expert talk about that, but we believe that is one of the most important components of the Med-ERRS process.

Also, because we give clinical information to our practitioners, I believe they're actually performing their own little failure modes while they're looking at the name. So having that clinical information, as Dr. Lesar mentioned previously, and looking at handwriting, they can look at it from the standpoint of filling an order in a hospital or a prescription in a drugstore to see where they think it could possibly be confused with

something already out on the market.

Another component that we think is really important is looking at the medication error literature to see what types of errors have already occurred out there so that even though we can't change those, we can be proactive with the trademarks that we're looking at and use similar types of situations that have occurred where errors have occurred to analyze the data that we're looking at.

And we actually will provide sometimes little snippets of medication errors that have been published in the literature as part of our final report to our clients so that they can understand where their product may actually be confused as well.

We do use computer searches, such as Thomson & Thomson. I think everybody here uses Thomson & Thomson, and also we do, because our practitioners or our expert panel -- excuse me -- are practitioners, they're pharmacists and nurses, that they are very familiar with drug information databases.

So we have on-line databases. When we sit around and do our failure mode, we have tons of books sitting on the table with us, the American Hospital Formulary System, the Red Book, the Orange Book, the Micromedex Facts and Comparisons. So we are constantly looking up increased information to what our practitioners have given us so that we can analyze the trademarks properly.

MR. LEE: Jim?

MR. DETTORE: And likewise we do a similar type of process. Our components include, as we said earlier, information from the physicians and the pharmacists based on interpretation studies as well as open ended input, and as Jerry said, it's important to bring forward the open ended information.

At the same time, we check a number of desk references. We subscribe to the National Drug Data File, which I'm sure any of the hospital attendees check on a daily basis. So we subscribe monthly to that, as well as 25 other reference

checks, some to include American Drug Index, Facts and Comparisons.

We're also checking the on-line ISMP, Michael Cohen and his fine group, as well as USP and ADI's Medication Errors or Confusions, and we check that. That's almost a must-do from any standpoint.

MR. LEE: Let me take you back and show you where we are on the box here. We've just been talking to the panelists about look alike/sound alike prescription testings and response to questionnaires.

Yes, Jerry. Sorry.

CAPT. PHILLIPS: If I can go over my process a bit.

MR. LEE: Sure.

CAPT. PHILLIPS: If I have that opportunity, and I'll let Toni also talk a little bit about the process at CBER because we don't have identical. We have a lot of similarities.

But the question was do we use computers and reference textbooks, and, yes, we do. We utilize Thomson & Thomson as was mentioned and all the

reference textbooks that are searched, the Orange Book and then the Red Book, and the PDR, et cetera, et cetera.

We also are developing a computer took that looks at the phonetic and orthographic similarity to trade names, and that should be up on line by October 1st.

And we also have expert panels which was mentioned earlier that incorporates the DMETS staff, along with a representative of DDMAC that looks at the promotional aspects of the name as part.

And then finally there's a risk assessment that's done by the safety evaluator.

So that's real brief, and I'll let Toni comment, too.

MS. STIFANO: Generally we follow the same steps with regard to having reference texts and on-line searches and the like to do an analysis of the sound alike/look alike promotional aspects and the like of the product.

Where we have significant problems, we

don't have the resources that my colleagues in CDER have. So where there are problems that are tough for us to resolve, we will, in fact, use them as a reference to take it through the process of the handwriting analyses and the phone-in and the like, and the other tests that they do where they can elicit responses that we're not able to do merely by using reference texts and on-line searches.

CAPT. PHILLIPS: And finally, if I could just add one more comment that I thought was important is looking at your post marketing experience, and we do tie in our experience that we have learned from errors that have occurred in the risk assessment of the trademark evaluation. So that's really a key thing for us, is to be able to do root cause analysis and learn from our previous experiences and apply that to pre-marketing.

MR. DETTORE: I'd like to add also that the Brand Institute additionally looks at the phonologic, syllabic differences in brand names as well sa bigram, trigram, and orthographic string

similarities. We do a number of the mechanical as well as reference books to complement each other.

Thank you.

MR. LEE: Clement?

MR. GALLUCCIO: Well, what I would add is that although we employ a similar approach, we actually introduce these methodologies prior to the actual creative development ever being shared with our clients because I think that, you know, certainly one lesson that we have learned -- I've been involved with this particular endeavor for, you know, close to 15 years -- is that we need to manage the expectations of what a pharmaceutical trademark should represent, and the earlier you can introduce the concept that one of the primary goals of this process is to develop a word that is differentiated and free from the risk of confusion and subsequent misprescription the better.

So we find that if you manage the process in that manner and introduce these methodologies very early on, the net result is that you have a group of

potential trademarks rather than an argument of whether or not you can secure that, quote, unquote, dreaded favorite name.

MS. STIFANO: One thing that I did forget to mention is that because there are a number of products at CBER that are not and under no circumstances would they be self-administered, something that people would pick up at a pharmacy, that they would be given more in a controlled environment, is that we have to look more closely at the elements of sound alike/look alike in terms of where they are on a shelf, how they're stored.

And so we take a slightly different tact with regard to worry about someone picking up the wrong prescription at the retail level. It's more of inadvertent mishaps at the hospital pharmacy or the doc.

And we have another confounding factor with vaccines, and that is as they become more complex and they start adding more antigens or whatever, what to do with names of things. Does it

change the product or not?

So that's something we have to start to address as things start to evolve in the area of vaccines.

MR. LEE: Just before we go to the final stage of questions, I just wanted to bring us back to this chart for a moment and what we've been trying to do is give you an idea of how the data is collected that forms the subset of names that we want to look at in more detail and determine whether or not there is a problem or not because that's the way the process is.

Of the universe of all names you test, you look, you try to form a subset of the closest names, and then from there you then have to make a decision and analyze these names and make a decision, and the last set of questions has to do with how this decision comes about after you form the subset through the testing and data collection phase.

So I'm going to direct this to Jerry and ask: Jerry, do you use an individual expert or is

there an expert committee? And I think you already mentioned that a little bit, but tell us how you evaluate the data.

CAPT. PHILLIPS: Well, I think we use both from my perspective. The individual expert is the safety evaluator who looks at the data, makes a risk benefit decision, analysis based upon the dosage form, the indications and usage, where it's going to be used, stored, et cetera, in order to reach an overall conclusion whether the name is acceptable or not acceptable.

So we do use an individual expert in that perspective. As I mentioned, we do have an expert panel. The expert or the medication staff who are attuned to medication errors, that's their day-to-day function. So they're quite attuned to it, and we also have a representative of DDMAC, as I mentioned before. As Dr. Jenkins was talking about the promotional aspects, we blend that in early. So DDMAC's opinion is actually part of the review that's forwarded to the Office of New Drugs, and that's part

of the expert panel that's done.

MS. STIFANO: Within CBER the review of, you know, potential proprietary names is handled by the Advertising and Promotional Labeling Branch, and there, again, it is very much folded into in terms of what are the promotional aspects of it as well as the potential for error in terms of sound alike and look alike.

And we do have the branch chief here, who is Glenn Byrd, and if you have any specific questions, you can direct them towards him.

MR. LEE: Clement?

MR. GALLUCCIO: We employ a similar construct. We have enjoyed an outstanding professional relationship with Dr. Neil M. Davis of Safe Medication Practices Consultants for close to 15 years. Neil has served as our independent to eyes and ears relative to the subject of medication error.

For those of you who have not had the pleasure, Neil is a co-founder for the Institute of Safe Medication Practices. It was in that role he

and Michael Cohen sat down with myself, David Wood, a number of others and said, you know, as a company that has a very large role in the development of pharmaceutical trademarks, there are a number of things that you really need to share with your clients and have them understand the dynamic of medication error.

So with Neil representing the expert, he has secured a number of individuals not only here in the United States, but worldwide that share this interest because they're not alone. I mean, they were certainly the pioneers, but as evidence here today, there are many individuals with an interest in minimizing medication error, and it's those individuals that we employ. They are practicing pharmacists, nurses, administrators in a hospital environment that assist in developing the safest name as we possibly can develop, recognizing the human element, the human factor.

MR. DETTORE: We are similar. We have our own outside expert panel, as I said earlier, but

we're really proud to have folks formerly from the FDA. Dr. Ben Lewis, 25 years, in the audience, as well as about 25 other of my directors from around the world are here today, but Dr. Ben Lewis who has helped at least Brand Institute understand the issues at hand, as well sa Nova's Southeastern Pharmacy School. A number of independent individual experts as well as on staff and outside nomenclature review folks help sort through these issues.

So I think we're all getting a feel there's no common one directional solution to prevent medication errors. It's quite complex, from the folks in the audience, Dr. Jenkins and every that has been presented before us. I think you're seeing a number of consultants here try to sort through these issues, and they continuously evolve.

I mean, this is going to continue to evolve, and all we can do is stay ahead of the game by having the best personnel on board to work with industry and the FDA in resolving these issues.

MR. LEE: Okay. Time is running short.

So yes or no answers on the next one. No.

(Laughter.)

MR. LEE: What scoring methods -- do you have objective measures or thresholds for establishing problematic name similarity?

This is a question I'd really like to hear the answer to. Jim.

MR. DETTORE: Excuse me? I'm sorry.

MR. LEE: Do you have objective measures or thresholds --

MR. DETTORE: Yes, yes.

MR. LEE: -- for establishing problematic name similarities?

MR. DETTORE: Yes, we do. Yes, we do. I'm sorry. You just want a yes. Yes.

(Laughter.)

MR. DETTORE: By the book, yes.

MR. LEE: You can say a little more.

MR. DETTORE: Yeah, we do have thresholds. We look at a number of variables. As Jerry said earlier, looking at an entire product

profile, everything from the classification indications, dosage strengths, dosage forms.

Oh, boy, my medication is kicking in. Talking about atenolol, 300 milligrams.

But we do look at this -- that's right. I went through an open heart surgery not too long ago. However, I'm not going to drop.

And with tha tin mind, we do a number of areas of assessing drugs, and we have thresholds for each of these.

Thank you.

MR. LEE: Clement.

MR. GALLUCCIO: I'll try to give you the short answer. I think that, you know, it certainly is a company who has been evaluating proposed pharmaceutical nomenclature. You do establish a set of benchmarks. I mean, we have been using the same model, so to speak, for roughly 12 years.

Now, prior to that most decisions relative to pharmaceutical trademarks was whether or not a physician preferred it over another name. But

since we've executed this model, we've conducted roughly 400 evaluations. So it's on the basis of those benchmarks that we have developed that we have the comparative context to make determinations relative to the ratio of individuals who correctly interpret the stimulus, those who misinterpret it, what they identify it for, and so on and so forth.

And beyond that, which was also touched on earlier, is that it's a fairly complex system of not only identifying whether or not the candidate can be correctly interpreted, but also once you had that misinterpretation, looking at those relevant dispensing factors, the route of administration, and so on and so forth. So it's all developed within a matrix so that with a weighted average you can have as best as we believe you can possibly have at this particular point, an understanding of the risk of misprescription.

MR. LEE: Scoring methods. Sue, use numbers, letters, words?

DR. PROULX: Well, we use all of those

depending on what our client wants, but if they don't have a particular preference, we use a one to five scoring range, one meaning that there's high vulnerability in our opinion of that trademark in the marketplace and five meaning that there's low vulnerability.

Not everyone wants a particular number. Not every client wants a particular number, particularly the trademark attorneys to tied a number to a score, but I want to also add that the number of the letter or the statement that we use to our clients, taking everything that we do into consideration to get that final number is really only for their internal use. It's not something that we want them to necessarily send to Jerry's group because those numbers don't mean anything to DMETS.

So we're giving them those numbers because normally we're testing, say, ten names. It's giving them a feel for where each of those individual trademarks that we've tested stands in our opinion from the safety standpoint.

MR. LEE: Jerry and Sue, I'm going to ask you the last question. Do you have waiting techniques for clinical variables when you do an assessment of trademark?

CAPT. PHILLIPS: Can I go back to scoring method?

MR. LEE: Sure.

CAPT. PHILLIPS: Well, the scoring method is problematic from my perspective. Our epidemiologists have had problems with trying to validate any type of scoring method there. So I don't use a scoring method. I just wanted to point out a difference in philosophy there.

If it could be objectively classified and validated, then that would be a good process to come out of this.

Do I have weighting techniques? I have a consistent process of looking at each trade name or trademark by a process using the expert panels of prescription drug studies. We consistently use the same process for every trademark evaluation, and we

do use clinical criteria in determining that, whether it's the dosage form. It is the dosage form. It's the indication, and there's outcome. What's the outcome of that particular error? Where is it going to be used and stored, et cetera?

And so, yes, we do.

MS. STIFANO: Ditto. In fact, we give, I think, a little more weight, if anything, to what the consequences are of a misuse or a mishap with the use of a product and the clinical elements of it in terms of where it is and how it's going to be used.

MR. LEE: Thank you.

DR. PROULX: It's a qualitative process so it's difficult, but we're trying to work on -- Metters (phonetic) is working on trying to come up with a mathematical way to weight thing. A lot of what Jerry said is very similar to how we do it, looking at all of that clinical criteria that we've been talking about the last 45 minutes or so.

We think one of the things that's important to weight perhaps differently with a

clinical product is that if it has a unique feature, for example, if it has a unique way of being administered or if it's given in a unique setting, then if something else is given in that way as well and there's confusion with it, that we would weight that particular clinical piece more heavily in that particular process. So we don't weight everything the same across the board.

Now, a pharmacist when you're looking at a prescription, the things that you're going to see are not necessarily all of the clinical information. You're looking at the name, the dose, the dosing schedule, maybe the name of the physician, et cetera.

And obviously as part of the process that we do, we're looking at those things as well, and we're actually still trying to validate ourselves whether the dose and dosing schedule are as important as we all say it is, and we're not quite sure yet.

MR. LEE: Well, we've moved to the panelists' remarks. We'll give each panelist about five minutes to talk a little bit more to tie things

together or express some factors that haven't been considered.

So I think you're up first, Jim.

MR. DETTORE: On behalf of Brand Institute, we wanted to thank, again, the FDA, Jerry Phillips, Michael Cohen of ISMP, and also PhRMA, Bob Lee, for inviting us here today.

I just want to take a few minutes about the collective mission, and as so many speakers today talked about it, it is a collective mission. It's a mission whereby, again, Jerry's vision in DMETS and ODS have given that direction to the manufacturers.

And the manufacturers, I applaud everyone here. The ones that we work with, and it's most of the majors and minors here, they are doing the due diligence up front on their research, and I think Jerry has had to pound it into everyone's minds that let's take it on our own to do this due diligence, and I don't see manufacturers at all seeing that Jerry's vision is not on target.

For research houses like Susan with Med-

ERRS up there and Interbrand Wood, I find colleagues here, as well as many of the other ones here. It's our job now to do the evaluation. You've heard of the number of different types of evaluation processes. I think you can probably get a feel for where you see the industry going right now, from there the health care community making sure that we get the checks and balances and the input from the health care community from not only monitoring our research, but also at the same time communicating to them the educational and promotional programs.

From there we have Michael Cohen, a fine group over there, ISP, as well as others, USP, all monitoring and surveiling and reporting on this.

You can see premarketing here going right around from right to left, and now you get to that surveillance area, and boom, an issue comes up. Now we go into the post marketing, and from the post marketing, it goes right back to due diligence, look at it from a risk management standpoint, taking those risk management strategies going to the health care

community, making a last check with Michael Cohen's fine group to see if it passes muster.

That's where the industry is going. That's where we all have to continue to go. This is a little bit about the road to our progress. It's been a pretty good road. It's been a road that 1997 in front of the LNC where Jerry and Dan Boring's group and Dan Boring's group and Yana Mille and a number of the other individuals that we've had personal one-on-one consults on behalf of the manufacturers themselves and their input.

We went from paper based on line, when Jerry took over ODRA, and I believe as a matter of fact Schering Plough is out in the audience today with Joel Wiener, and that was my first time I went into the FDA being six, two and you came out three feet, and you learn a little bit and you become a lot more humble and at the same time smarter.

And from there we took it to the ODS and the DMETS of 2001, and what's interesting here, we went from now a one, a unidirectional of one faceted

approach to a multi-faceted approach, including qualitative, quantitative, your own expert panels, and at the same time our own internal folks.

An entire staff headed up by Dr. Kovara (phonetic) down in Miami has been working on a number of names, and you're going to see here some of the output. The output in our methodology premarketing and post marketing, making sure that we identify the risk up front, the interpretation studies via the docs and pharms., as well as on-duty prescription studies.

Next, the assessment. That's where we go into the various research reference checks, as well as regulatory guidance review, and also a professional review committee externally and internally.

At the same time now communicating that to our clients, and doing the due diligence on behalf of consultants to make sure that that information in any types of issues with errors are communicated to our -- and any kind of nuances that we're learning

from personal consults.

Our clients have been very open to us. We're submitting now two or three white papers each week on behalf of our clients to the FDA, and going overseas it's the same way to the EMEA. The FDA is very, very open to research from the actual consultants and, most importantly, the sponsors. So we're having a very good success rate at this point.

Post marketing risk minimization, that's why we have Dr. Lewis, Dr. Carsten, and a number of outsiders as well as insiders looking at risk benefit management programs based on product labeling, product packaging, and a number of other areas there. The best then go back for corrective measurements.

From this I have to, one, state and I have to applaud the industry. It's been a tough road. We do a lot of naming with eight offices out there just strictly focused on medication error evaluation, and we are one of the top ones, as we have here, that review our names.

This is an interesting chart. I want to

take a few minutes. We have not randomly selected. We have taken over 1,500 of our names in the last six years from that same chronology of the old LNC under Dan Boring and Jerry Phillips to when ODS came about and then obviously most recently DMETS.

And we have taken our second and third and fourth generation models and then reviewed that based on the currently marketed products that have been approved by the FDA, and those citations based on ISMP, USP and ADI, and I'm not here to tout Brain Institute. I'm here to tout a process.

I'd like to say we're really close to the industry. We get right in front of our clients. We're there to help them, as well as the FDA, and we take the knocks from the FDA. We feel pretty bad, too, when we get a rejection, but at the same time, it's a learning process, and I think we all have to understand it's a learning process and we'll continue to near. This zero will not stay zero, but I'm sure we're going to continue to do everything possible to make sure that our methodology continues to move up

the ladder and up the standards based on the entire industry's focus. We have to be able to evolve. We have to be able to react, and at the same time, after that product the product quality.

And I applaud the FDA, and Jerry has been pounding it in the audiences' heads a long time. Let's start working towards a common goal: patient safety.

And these numbers are showing it. These are our actual numbers from Brain Institute. It's not the industry. It's us. So there were 1,500, 1,513 names tested during that six-year period of time, and I'm very proud to say that we are starting to improve as an industry towards our common goal, patient safety and monitoring.

Thank you.

(Applause.)

MR. GALLUCCIO: Once again, thank you to all for putting together this very informative and much needed discussion about the challenges that we face in developing pharmaceutical nomenclature.

I only have this one slide to share with you because it represents what I would define as the best practices, and we've used a number of key words here today, methods, approaches, processes, and so forth, and I have incorporated all of them within this slide.

But to be completely honest with you, I personally view this as a philosophical question. Will you, as the sponsor, use all resources within your power to develop the best possible trademark from a patient safety standpoint?

And one of the classic definitions of insanity is to repeat the same behavior over and over again and expect a different result, and the fact is you can have the most robust validation process, but from a philosophical standpoint if you are developing names that are perceptually similar to presently marketed trademarks, you will find yourself in that endless cycle of rejection and more evaluation and more rejection.

So there are three steps as I see them to

implementing a set of best practices within your organizations or perhaps sharing them with your clients.

The first is to influence whoever is involved in this particular endeavor the importance of a safe trademark, and what we have implemented at Interbrand Wood Health Care in our Rx mark is a series of screens known as conflict filter, and one of them employs a computer assisted decision analysis tool that provides metrics relative to the similarity of a candidate and a presently marketed drug.

So with this pre-screen, we actually eliminate close to 70, to 80 percent of all of the creative that's developed for a particular requirement.

The client only has the opportunity to see those names that should be highly differentiated. Are there instances where perhaps we pushed the envelope a little bit? Yes, that's true, but fundamentally what is provided for the process begins with a very clean set of potential candidates.

The second step is related to the methodologies that shared with you here today, and certainly I believe that if I was sitting in the audience, and I was listening to the panel, there's no questions that there's a great deal of overlap and similarity with our approaches, and that's for good reason. I believe that we all have a fairly good sense not only what may be the best methodology in an academic context, but what also would provide that final deliverable, something that a team can agree to.

So I think the guidance there is that your assessment of a potential trademark should reflect multiple data sets, qualitative as well as quantitative, an expert or an expert committee, as well as all of the wonderful tools that exist today that did not exist ten years ago. Electronic tools are just adding so much value to the identification of potential conflicts.

And finally, something that we've recently begun to explore and implement for our

clients is the concept of a dispensing advisory board. Identifying individuals internally as well as externally, and although I've listed this as Step 3, this could actually be the very first step, and implement their guidance within your own internal SOP and have this advisory board confer at all of the significant milestones within this process up to and beyond launch.

And I believe that you will find by creating this dialogue, being open to new ideas and recognizing the fact that the very best trademarks do not sound like everything else that's already out there, but are differentiated, innovative, and are protectable from a legal standpoint, I think that that would certainly give you the type of record that we have enjoyed over the past 15 years.

Thank you.

(Applause.)

DR. PROULX: Thank you.

I think this has been a good start to the day. I hope that we've stimulated a lot of you and

that we get some good questions after our presentations.

I just have two quick slides, and I really want to reiterate some of the points that came out when I was answering the questions.

This is the Med-ERRS process, quick and simple. We have a niche. We're doing safety testing, only we're not involved in name development. I'm sure we've seen some of our colleagues' here names and done some of the safety testing on them over the years. So we're coming from the standpoint of only testing the safety of the names at Med-ERRS.

One thing that you can help us with as clients is to give us enough clinical information, and we talked about where in the process of drug development you should be giving us the names to test.

And I have a very difficult time trying to explain to clients that we need clinical information so that if you're not giving us a product that's far along in development without a dose,

without a route of administration, without the dosing schedule, it's very hard to come back with good information on whether the product is going to be confused in a real life setting. So you can help us with that.

That's the first thing that we need, and that's the product information.

Project coordination is an internal step that we use at Med-ERRS just to get things going. I know we all have our own different process, but, again, there's a lot of similarities as well developing the data collection tool and notifying the practitioners who are appropriate for that particular product to test, as well as who would be perhaps dispensing that product in their clinical site.

But the two that you see I have highlighted are practitioner input in Med-ERRS analysis, which is what we consider our expert panel. As I said earlier, we believe that the practitioners are doing their own failure mode in effects analysis at their sites when we're giving them the appropriate

clinical information, handwritten and verbal information about the names, and that's important. That's a good step to start.

And what we do is then take that information back to Med-ERRS and do a more comprehensive failure mode and effects analysis with our own internal experts.

And I didn't say before, but I believe that we should use a panel, that there should be some consensus to reaching decisions on each final trademark that we're testing, and using that other important information that we mentioned before, the computer searches, the drug information literature, et cetera, when we can finally give you a good final report, a good assessment of what we think of each trademark.

So really, I think I just said what I need to say. The practitioners are important. Taking that information and prioritizing that, letting the experts do that; use of failure mode and effects analysis, which I think we'll hear about a

little bit more this afternoon.

I can't stress enough look at the literature. It's the same idea as what Clement just said about making the same mistakes over and over again. You don't want to do that, so that if you can learn -- we tell this to our practitioners that we talk to day in and day out. Learn from the mistakes of others. So look at the medication error literature, and we try to take that and use it specifically toward trademark testing.

But as I said before, it's a qualitative process, and I think it's going to continue to be at least partly qualitative.

So I really do look forward to the experts this afternoon who are going to be talking about the various components that we discuss to give us their insight, and we have always striven to do a better job and to improve our process continually, and we hope to continue to do that through this meeting today.

Thank you.

(Applause.)

CAPT. PHILLIPS: Okay. I think we're getting close on my time here. So I'm going to go real fast.

This is just basically an overview of what you've already heard today, just to go over the process a bit.

We do begin our review at the end of Phase II of an IND, and we also perform another review. It's an abbreviated review that's done 90 days prior to approval. The objective of that secondary review is to look at trade names that were approved by FDA from the time of the first review until the approval of the NDA.

As mentioned we do a proprietary name analysis, which exists of an expert panel review; verbal and handwritten prescription studies; and a computer assisted analysis.

We haven't focused much on this, but DMETS also looks at the labeling and packaging, looks at the container, the carton, the package insert,

proposed packaging configuration, whether this is going to be a syringe, ampule, an oral tablet, et cetera.

We look at the overall risk and benefit in order to make a final evaluation. Written recommendations are then provided to the reviewing division, who consult at the Division of Medication Errors and Technical Support.

As mentioned before, we are looking primarily from a safety perspective for sound alike/look alike names, the currently marketed drug names to other medicinal products, to medical abbreviations, procedures, lab tests, et cetera.

Contributing factors, we mentioned this, are similar indications, although just personally having two products that have different indications, I don't consider that a very powerful reason to say that the two names should be allowed, mainly because practitioners really don't know the indications for a patient when it's dispensed for the most part.

So the opportunity for error will occur

irrespective of indications. The same patient population are contributing factors, identical formulations, overlapping strengths or directions, similar names that have identical strengths and identical routes of dosage administration -- they're administered the same -- are pretty much rejected for the most part. It's a pretty strong criteria.

If they're stored in the same area, that increases the risk, and this is a process. Just to give you an overview, the name comes in on the left-hand side from the product sponsor. That's the blue box up there. It goes to a project manager within the NDA or the IND that you file.

They, the project manager, will consult DMETS to the right where we do an analysis. That analysis, as I mentioned, computer analysis expert panel, an Rx study. That's all coordinated by a project manager within DMETS.

That name is then sent to a safety evaluator, who looks at the data from those previous studies and expert panels, puts a risk benefit

analysis together. It is routed to a team leader and the Deputy Division Director of DMETS, then sent to the Office of Drug Safety for final review and signed off and is sent back to the reviewing division.

And then the reviewing division is responsible for notifying the product sponsor whether the name is acceptable or not acceptable.

So with that I think I will stop. Thanks.

(Applause.)

MR. LEE: We'll just wrap up a little bit, and then we'll have some questions.

I think the real issue is what we're trying to do is look at these systems, and this afternoon we're going to be listening to some experts helping us with this because the goal really is to get information which is reliable and information that is really relevant and try to avoid last minute, subjective judgments about name availability because they can be very disruptive to bringing the product to the marketplace.

So we'll listen to a couple of questions if there are any from the floor.

MR. DOUROS: I appreciate the point earlier that any feedback from the agency, say, at the end of Phase II is preliminary in nature, and I understand the reasoning. Can you tell me if the agency is looking into any initiatives or alternatives to work with sponsors a little more proactively to sort of avoid the result of having a yes/no decision 90 days or less before approval, you know, a similar initiative maybe to the TPI initiative?

CAPT. PHILLIPS: There is a discussion. There's a draft guidance document that I've been working on for a couple of years.

(Laughter.)

CAPT. PHILLIPS: You know how it takes a long time to get out, but in that guidance document, and we have had discussions in the center with PhRMA on this issue to try to improve the process, the transparency of the unapproved names that are in the

pipeline.

And the proposal that we've discussed is to, after the tentative decision on a trademarked is found acceptable, with the agreement of the sponsor, the FDA would put this onto the Internet so it would be visible.

It would be the name of the product, the proprietary name, the name of the applicant, and the date it was found tentatively acceptable. And what that would do is let you see what the acceptable names are that DMETS or the FDA has finally agreed to. That's tentative.

Is that --

MR. DOUROS: Yeah, I think that goes some part of the way to providing the information, but that may also cause further disputes among sponsors. Is the agency going to play a role to work with, you know, different sponsors at various points in the process to, you know, bring about more certainty?

CAPT. PHILLIPS: Now, there's been a couple of occasions where we've had anticipated train

wrecks, and we know what's coming through the pipeline. So, you know, we have confidentiality issues from one sponsor to another not to acknowledge an application. So the way I've worked with the reviewing division is to notify both project managers in those review divisions that the situation is one we'll have to change their name based upon the approval, and to ask those project managers to notify the applicant holders to work, to agree to work together.

And the results, you know, are whoever has -- I would say if you have a priority review and you know you're going to get your application approved first, there's probably not a good reason to cooperative, but I at least offer that opportunity for the conflict to work itself out through the two sponsors and not FDA getting involved with that.

MS. STIFANO: I think Bob Lee can speak to negotiations at the nth hour where we have tried very hard to avoid -- well, tried very hard to resolve a situation that seemed difficult.

So, yes, we will work with you.

MR. LEE: Yeah, I have to say we were very familiar with the CDER process, but we weren't as familiar with the CBER process, not having as many CBER products, and so we has a problem with the trademark on going through CBER, and very, very open communications with Tony who worked. We tried and tried and tried to get our original mark through, had a big splash in the Wall Street Journal. It was very important to us.

But at the end of the day, we just couldn't work thing out. It was certainly not an unreasonable problem that the FDA had. So we went to another mark, and we were able to get that second mark approved very quickly. I think the agency tries to cooperate when it can.

MR. DOUROS: The other question I had for the panel in general, I think some of the most recent numbers I saw was attributing medication errors to -- 15 percent medication errors attributed to confusion among names. Is there any sense as to how much that

can be improved simply by trying to reduce confusion in the names?

I recognize that the numbers are probably a little fuzzy, to begin with, but how much improvement do you think is possible here? Do you have a sense of what we're looking at?

Thank you.

DR. PROULX: The numbers I've seen are actually closer to 25 percent. The 50 percent is related to labeling, packaging and nomenclature, which are all issues that the industry and the FDA can get involved in.

I'm not sure I can answer. Mike Cohen might be able to answer that a little bit better, but we know that products have been confused due to their similar names even when no other clinical factor has been similar. So that's why before I had made a comment about we're looking at the fact that is does that important.

And it is in a lot of cases, but when you analyzes these errors, you can have two products that

just have similar names that are given by different routes, that have different dosage chains, different schedules, and through the drug use process somehow one gets confused for the other.

I don't have numbers, and as I said, Mike is going to be up later. So maybe he'd be better able to address that.

And there's Mike.

DR. COHEN: Thank you.

Well, looking at the data from the medication errors reporting program, I think Susan is closer to right at about a quarter being name related. However, that's certainly not just brand name. It's nonproprietary name, and certainly there are lots of reasons as was pointed out before that contribute to those errors. So that's something to keep in mind.

Some of it is the suffix situation that we talked about, for example, but I think what you have to keep in mind is the reports that are received at FDA and ISMP-USP's program come from practitioners

who are concerned about the product more than practices, and I think that kind of skews the data that we get.

We're likely to get product related medication errors through those programs. So it really bumps up those figures even higher than they actually are.

The only way to find out for sure would be to look at actual data reported within a hospital. Large databases, for example, exist in this country, or to do direct observation of medication administration or other studies that have been mapped out for ambulatory care. So we don't really have a good handle on that figure, but it's nowhere near 50 percent. I agree.

MR. LEE: Mike, before you leave, do you think that as we would move away to a new prescribing and dispensing environment, like E-prescribing, and you would avoid handwriting, you would avoid verbal ordering which often are factors in making --

DR. COHEN: Yes.

MR. LEE: -- two names which are otherwise acceptable unacceptable, that might improve as well then, right?

DR. COHEN: Absolutely. There is no doubt that electronic prescribing, electronic transmission will vastly reduce these communications errors, the look alike/sound alike issues.

It's still certainly possible to choose wrong items off of a computer screen, which we have seen happen many times. You could still have look alike drugs. I think some of the things that have been done to prevent that, like using tall-man letters, have been very helpful.

So that's something we should continue to do, but no question in my mind errors will be reduced, but there will be other problems that come with the technology, too.

MR. LEE: I have one back here first, Bruce. If we could take the one in the back first.

Thank you for coming to the mic.

PARTICIPANT: I know our focus here is on

the proprietary names, but I had a question, I guess, to Jerry and Toni on what, if anything, the FDA is doing now about the nonproprietary names and confusion in that regard. Is there a formal process that's being used to evaluate that at the time of approval, or is it just basically taking, you know, a formulary name, like the USAN or a compendium name, and sort of accepting that?

CAPT. PHILLIPS: We don't limit our review for looking at proprietary names just to other proprietary names. So we will look at confusion potential between a trademark and a generic or an established name. So we'll look across in both of them.

And because we have the opportunity to see those names in the IND, there will be an opportunity and we have commented on the similarity of the established or the generic names to other established names and have recommended that, you know, the sponsor be notified of that.

Of course, the established name is given

by USAN and is not an FDA responsibility, but being early involved in the IND process gives us an opportunity to point out similarities and try to get those names changed before it's marketed, and it's not an easy thing to change an international name.

DR. PROULX: Some clients have actually requested testing of nonproprietary names as well that we've done for them.

MR. LEE: May I ask that anybody asking questions identify themselves before they go forward.

DR. LAMBERT: Bruce Lambert from the University of Illinois, College of Pharmacy.

The FDA's job, it seems to me, is to balance risk and benefit for the public. So even drugs that are plainly toxic like thalidomide or chemotherapy agents can get approved because the benefit outweighs the risk.

But in the context of drug names, what's the benefit? To the public what's the benefit?

We see what the risk is. The risk is name confusion, potential patient harm, and so on,

but what's the benefit of a drug name that would counterbalance the risk of confusion?

It seems to me or I can't think of the benefit to the public. There are benefits to the sponsor of people like the name, they remember the name, it might make them prescribe the product, but I don't understand what the benefit would be to the public that would counterbalance the risk of confusion.

MR. LEE: Well, I think a trademark is really a two-sided coin. There is a benefit to the owner of the trademark, but really that benefit is because of the perception that the public gets when it sees that single name, the trademark. It can rely on that name to provide the collective quality and experience that the patients had with that product over the years in a single name. They know that that product will give them that same quality time after time.

And I think that it's not something easily measured, but they will often say Coca-Cola is

the big asset of that company, and it makes billions of dollars every year and it's all wrapped up in a single name.

My daughter takes an anti-smoking patch, for example, and she's had both the brand and alternatives, and I don't know whether it's the acrylic adhesive or what it is in the patch, but she has repeatedly tried to go off the brand and found herself going back to the brand because she gets better performance in terms of eliminating rashes and things of that nature.

So the brand name says this is the kind of quality, the kind of performance you're going to get time after time out of that product. If you have a generic name and it comes from a variety of different sources, a generic product -- I mean this is in any industry -- facial tissues, you don't know what ply, the number of plies you're getting, how many of these facial tissues you're going to have to use to do the same job as one facial tissue from a brand.

So, I mean, I think that's the kind of performance that you can rely on, and it's not necessarily high quality or low quality. It's a question of the value that that product is affording you, the quality for the price.

DR. LAMBERT: But that's the benefit of brands in general. What's the benefit of a particular brand which would say, "Well, we'll accept this name even though it has this risk because this name has some benefits"?

When we're balancing risks and benefits with drugs, we say we'll accept this drug because this particular drug has benefits which counterbalance its risks. When we accept a particular brand, there's always an alternative name, and so this gets to this issue of zero tolerance which somebody brought up earlier this morning.

I don't know the answer to the question myself, but I think it needs to be asked. How much risk should we accept in a brand name when it's not balanced by any obvious benefit to the public?

MR. LEE: Well, we can debate that,a nd we don't have the time. What I can say is that you have to identify the product somehow, and so any identifier you put on the product can be mixed up in the market place. Brand names certainly undergo a rigorous process before they're selected.

Let me mention a couple. We don't have time for anymore questions unfortunately. Sorry. Trying to keep on schedule.

But I think two notes here. One, the break starts now. Fifteen minutes. We're trying to keep on schedule. So please come back in 15 minutes.

And those who are speaking at the public session, would they please assemble right after this here at the front of the room? The speakers do that before the break.

(Whereupon, the foregoing matter went off the record at 10:35 a.m. and went back on the record at 10:51 a.m.)

CAPT. PHILLIPS: All right. We're going to get started now.

All right. In case you don't know who I am, I'm Jerry Phillips. I'm the Director of the Division of Medication Errors and Technical Support in the Office of Drug Safe and the Center, and I'm here with Toni Stifano of the Center for Biologics, and we are here to open the public discussion this morning.

The questions that were posed by Dr. Seligman were given to you. These are the questions that we will be listening for public input. Each speaker has seven minutes to talk, and I'll be kind of watching that seven minutes and giving you a heads up if you've gone over, and we'd appreciate it if the speakers would stick to a schedule of seven minutes.

The public docket is open for public comment. It is open until July 15th, and if I could, I wanted to make a comment that relates to the overall discussion here today. On our last panel we had certain companies that were part of the panel. I would like to let you know that this was a representative sample of an industry. There is no

way that there is an endorsement of any one particular company. There are plenty of companies that do this work.

So with that, I will open this for Susan Winckler. Susan is the Vice President of Policy and Communication, staff counsel at the American Pharmacists Association or Pharmaceutical Association? Pharmacists. I was right the first time.

MS. WINCKLER: Good morning. Thank you for the opportunity to present the views of the American Pharmacists Association.

APHA was founded in 1852 as the American Pharmaceutical Association, and we only changed the name April 2nd. So it's okay, Jerry. That's a name change that we're trying to get everyone to agree with.

APHA is the first established and the largest national association of pharmacists in the United States. Our 50,000 members include practicing pharmacists, scientists, student pharmacists and

pharmacy technicians, and obviously decreasing medication error and improving patient safety is an essential element of what our members do.

The similarity between drug names that sound or look like the names of other medical products has been identified as the source or at least a contributing factor of many medication errors. While we do not know how many medical mistake are directly attributed to sound alike or look alike drugs, approximately 25 percent of all medication errors reported to the USP medication error reporting system are due to similarity in drug names.

A recently published study in the Journal of the American Pharmacists Association by Professors Barker and Flynn also noted an incidence of errors in sound alike drug names.

These are frightening statistics, and the number will grow if we don't employ a systematic approach. It will grow because we have a number of new drugs entering the market, a number of new drugs

that pharmacists and physicians and other prescribers and consumers must manage and understand.

Each of the new drugs must have a new name, and it's becoming harder and harder for manufacturers to develop new names that are both short and catchy to meet their marketing concerns, and more importantly, unique and that don't conflict or sound like other medications.

We're pleased that PhRMA, ISMP and the FDA convened today's meeting. Any effort to decrease confusion related to drug names is a welcome step.

While we do not claim to have the specific solution to this problem, we will offer the following three thoughts for your consideration:

The need for guidelines or consistency for evaluating names;

A support for reviewing both prescription and over-the-counter names;

As well as some recommendations for the studies and evaluation that are done on those names.

One of the questions posed for this

meeting concerns the current methods employed by drug sponsors and the FDA to evaluate drug names. As we understand the current system, there is no consistent method of name development or evaluation currently in use.

Historically sponsors of proprietary drugs developed a drug name and submitted it to the FDA for consideration. In the past few years, manufacturers of proprietary drug products began conducting their own name studies.

While this frees the agency from conducting naming studies of its own, it raises concern about the consistency of methods used to identify concerns with those drug names.

As the FDA looks to address the need for consistency, we support a concept that I believe was termed the good naming practices this morning and suggest that that something that should be strongly considered.

Additionally, as a good naming practice or any other type of system is considered, we also

suggest that you look at other systems, specifically the drug naming process for non-proprietary names.

The United States Adoptive Names Council, of which APHA is a supporting organization, has specific guidelines for assigning generic names. Before the USAN council will approve the generic name, it follows a process to insure that the drug name is appropriate for the product and that it is not too similar to an already existing name.

While the USAN method is not foolproof, as no system is, the system relies on a standardized process. We recommend that the agency and the industry examine the USAN process and adopt a more systematic process with standardized tools to develop and evaluate drug names for proprietary drugs.

Another question for today's meeting concerns evaluation procedures for different types of drug classes, such as prescription and over-the-counter medications. We feel strongly that drug name safety testing for all medications, regardless of their class, should be held to the same high

standards.

Eliminating confusing nomenclature practices for all medication products is an important step towards reducing medication errors of all kinds. As Dr. Jenkins noted this morning, there's a particular concern in the OTC category, and that's the family name concept or, as we call it, the brand name line extension.

In the OTC environment where you have brand name line extensions, where the same brand name is used for a number of different products, something very important happens. We all assume that with a certain brand name you will get a medication that has a certain active ingredient, and that is not true.

Many consumers do not use the full name when they're referring to their OTC products. So if they refer simply to the shortened brand name or brand name without a suffix, the pharmacists and physicians who are trying to work with that patient really don't know what the consumer is taking, and that obviously creates a challenge to our mission to

improve medication use and advance patient care.

The last question I will address concerns the kind of information that should be included in oral and handwritten prescription drug studies. This is a difficult question.

In an ideal world, prescriptions and medication orders would be typed or transmitted electronically. They would include all relevant information, such as the drug name, strength, quality, patient directions, and indication for use. If that reflected a realistic prescribing environment, it would be inappropriate to include all of that information in the drug name studies.

But we don't live in an ideal world. In reality, prescriptions are often transmitted orally. The majority of paper prescriptions are handwritten, and many are hard to read. Many prescriptions do not contain all of the relevant information, and on occasion prescriptions arrive with the drug product's name misspelled.

This reality needs to be considered when

designing drug naming tests. In order to assess the potential for a name confusion in a real practice environment, a number of tests should be conducted that include a minimum of information and in some cases perhaps a misleading drug information.

A health care practitioner is more likely to select the wrong medication when the drug product's name is misspelled or when the information available to them is minimal. An example of a confusing drug pair is Celebrex and Cerebrex. They sound the same when transmitted to the pharmacy over the phone. If the name of the drug is the only information that the hospital pharmacist receives, then the opportunity for drug name confusion is high.

However, is a prescription drug order includes additional relevant information, such as the route of administration, the nonproprietary name or the intended use, the opportunity for a medication error decreases dramatically.

There's one piece here that's also not addressed in the questions that were posed. The

questions specifically asked what we should look at when we're testing handwritten and verbal or oral prescription drug orders, which seems to presume that if we have typed or electronically transmitted orders that there won't be a potential for confusion with look alike drug product names.

I don't know that that assumption or presumption has been proven and would observe that it's something we need to look at. As Dr. Lesar noted this morning, they have found problems with the technology they've used with prescriptions that, indeed, you do still have confusion in that environment.

And I think all of us have numerous examples where in our word processing or in our PowerPoint presentations we don't have anymore misspelled words, but we have correctly spelled wrong words. Correctly spelled wrong words in the prescribing world gives us correctly spelled wrong drugs and is something that we need to avoid.

And if there is a way to assess that in

these studies, we should look out for that.

In conclusion, I would like to reiterate our support for the activities of the groups gathered here today. Measures to decrease medication errors and increase patient safety are a top priority for APHA and our members. With confusion over look alike and sound alike drug names responsible for a significant portion of medication errors, the development of a standardized evaluation system that makes use of standardized tools is critical to improved patient safety.

Each drug should be extensively examined for any similarity to an existing product and evaluated as it would be used in a real practice environment. While developing a name for a drug is driven by many different factors, the primary measure for evaluating a name must be safety.

Thank you for your consideration of the views of the nation's pharmacists.

(Applause.)

MS. STIFANO: Our next speaker is Maury

Tepper, III, from Womble, Carlyle, Sandbridge and Rice.

MR. TEPPER: Thank you.

And I somehow knew I was going to need to adjust this microphone.

(Laughter.)

MR. TEPPER: I do want to add my thanks to those that you've all heard this morning. this is in many ways a historic meeting, and it's a pleasure to see all of the different interests represented here working for a common goal because, indeed, that's what we've been doing for years. We may all have some different perspectives on how we ought to go about this, what the best way is, but clearly we're all interested in patient safety, and we need to do everything we can to minimize medication errors.

And I think that the effort being put forth here is a testament to that. So I applaud that effort.

I would like to draw out and propose that

a lot of what we are doing here actually is not a departure. I hope you've heard this morning from some of the companies presenting the lengths they go to in their trademark evaluation and clearance process, the way that these new analytical tools have been implemented.

And I would actually propose that these fit very nicely into existing legal constructs that we've had in place for years and that the decision making can most appropriately be made by turning to trademark laws well established likelihood of confusion standard.

I was struck, and I think hopefully you were, by Bob Lee's charts when we looked at the panel earlier today on the process, and if you looked at the legal clearance, it was all about data assembly, identifying a subset of potential conflicts, and then conducting an analysis and decision making.

And when we went to the medication error testing tree, it was the exact same process. Where we need to focus our attention is on that decision

making point. How should we conduct the analysis and how best can it be done?

I think we all agree that collecting this data is important. But I think we may be overlooking that we are basically trying to accomplish what trademark law does. It analyzes the similarity of marks and tried to identify the potential or likelihood for confusion from the perspective of consumers in the relevant marketplace in the way they encounter these products.

Now, what is unique about the prescription marketplace is the way in which these marks are encountered and dispensed. It's about the only place I know of where the consumer, the ultimate consumer of the product doesn't make the purchasing decision and is not involved in the selection and dispensing.

That gives us some very particular circumstances we need to look at, and indeed, the analyses that we see here about trying to identify handwriting peculiarities, similarity in

pronunciation, dosing strengths, and indications in many ways go to the fact that often the consumer walks into a pharmacy, never read this prescription, takes home a bag, didn't look into it, and does not play the traditional role that one would in brand selection and reliance.

So we need to be analyzing proposed trademarks in a way that take into account how it's encountered in the marketplace. Our legal system and analysis is set up to do that. We have an agency that has more than 100 years of experience in applying these tests and refining them, and we have a predictable set of rules.

These analyses actually fit very nicely into that. Trademarks are all about establishing unique brand identifiers that people can recognize, that can readily be distinguished from others, and basically there's consumer protection at the end of the day. That's what we're working for here.

I think that try though we may, and I heard it come out several times today, the end

decision cannot be mechanical. I wish we could have a formula. I wish we could have a score. I wish we could have a number that would absolutely tell us will there ever be an error. Will there be a zero level of errors, and, Bruce, you know, your question about what is the benefit, how do we decide when the additional risk of some error, you know, justifies going forward?

This is a difficult question. At the end of the day, this is going to be a subjective call, folks.

I applaud the efforts to look at algorithms, analyses, formulas to try to do it. There are simply too many factors that have to go into making a determination of errors given the many conditions in which these products are dispensed, countered, different dosage strengths, the ways in which the prescriptions may be written.

And what we've heard today are different ways of taking that all into account, but then someone has got to sit down and apply some useful set

of analysis to arrive at a decision, and trademark lawyers have been trained to do that. The legal system has allowed for review of that application of that standard, determination of priority and rights, and we have a clear set of analysis and guidelines to work from there.

We should look at agreeing upon this whole set of data inputs. We should continue to work here to agree upon the appropriate tests and analyses to be conducted, but then recognizing that we have a subjective decision to be made, we should turn to the legal system that has extensive expertise and has developed the factors to apply.

We should incorporate the data input into that analytical method, and we should arrive at a decision.

We should also be very cautious then, once we've assured that all of the proper steps have been followed, that all of the analysis has been taken into account to substitute one subjective decision for another.

Unfortunately, our current system involves that. Everybody is trying to make a determination. The sponsor has done, as you can see, extensive testing of the same sorts and arrives at, you know, hopefully, a reasoned and subjective decision. FDA does the same type of analysis and arrives at, again, a very well reasoned and subjective decision.

We need to focus more on agreeing upon the process. We need to focus more on insuring all of the steps have been followed and expend our energies there.

Lastly, I want to put this whole issue into context. I think it's an outstanding effort. I applaud all of the parties here involved for the work they've put into this and will continue to do, and I hope we won't lose sight of the role that this plays in an overall systemic approach to reducing medication errors.

The statistics that we've heard are certainly serious. It's difficult to ascribe though.

The 26 percent, is that the name causing those errors? I think all would agree any time an error happens it's the result of multiple failures in the system. A lot has to have gone wrong, more than just the fact that we have similar names.

And I think we want to continue to work to rigorously analyze these name to come up with a way of predicting and avoiding that similarity whenever possible, but we ought not lose sight of all of the other efforts and impact we can have on looking at the system, looking at the way in which indications are given, looking at the way in which product names are written out, looking at how dosing strengths are questioned, educating patients to take part in this process and ask questions about their medication so that we can have a bigger overall impact on the reduction of medication errors.

I would love to see a zero level. I think we all know that we'll get somewhere close, but not there. But I think that we need to also keep in mind where we can have the biggest impact on that

system and continue to work on this issue, but not lose sight of the many other factors that are having an impact on this problem and do what we can to address those as well.

Thank you.

(Applause.)

CAPT. PHILLIPS: Thank you, Maury.

The next speaker will be Dr. Bruce Lambert, Associate Professor at the College of Pharmacy at the University of Illinois.

DR. LAMBERT: Thank you for the opportunity to be here today. I appreciate it, Jerry and Mike Cohen and Bob Lee.

I'm going to read some of my remarks. It will be available eventually when we submit our testimony into the docket, but if I read it, it will go more quickly than if I extemporize.

So for Question 1, the following considerations are important when evaluating any proposed method of evaluating trademark names.

One, the method must be scientifically

validated. That is, there must be some peer reviewed evidence preferably that the method being used can actually reduce the probability of confusion or harm. Ideally this validation would be based on some form of behavioral test of memory, perception or action conducted under realistic circumstances.

(b) The method must be reproducible and to the extent possible transparent. That is, others must be able to clearly understand and independently reproduce the result of an evaluation. This may be difficult given the place for safety screening services and the related need or desire to keep some methods as trade secrets.

(c) The method should be at least in part objective. Although the subjective judgments of experts will inevitably be relied upon in the final analysis, as Maury said and I agree, we would never consider making safety or toxicity judgments in the absence of objective data, and we should not make naming decisions without objective evidence either.

(d) The circumstances of evaluation

should be free from real or apparent conflicts of interest. One potential source of conflict that needs to be dealt with is when the organization who coins the name is also the organization that screens the name for safety.

If an organization has a financial interest in the eventual adoption of the name, some safeguards must be put in place to try and make sure that those who would benefit financially from the adoption of the name do not unduly influence the safety screening of the name. This might be done by blinding or by other mechanisms.

I should note that obviously in the pharmaceutical industry the companies who sponsor the drugs also do all of the clinical testing of the drugs. So we're not going to avoid this completely, but I think some thought needs to be give to this issue of real or apparent conflicts of interest in safety screening.

(e) In normal FDA safety decisions, and this goes to my question before, approval often

hinges on risk benefit analysis. If the benefits outweigh the risks, then the product is approved for sale. In the realm of drug names the risks are fairly clear, but the benefits are not. The risks are. When drugs are confused, patients can be harmed or even killed, even though having said that, we should note that most errors do not cause harm thankfully.

In the context of drug naming, however, the whole notion of benefits is not clear. What are the benefits of a drug name that might justify accepting some level of risk related to confusion? There are no clinical benefits of one name over another, are there?

So one must conclude that the benefits are commercial, i.e., marketing benefits that accrue to the firm who manufactures the product. The benefit of a good name is that people like it, have a favorable impression of the product, remember its indication, maybe more likely to prescribe or request it.

These are not benefits that accrue to the public. They're private or corporate benefits, but the risks, on the other hand, accrue primarily to the public. So a fundamental question is whether we should trade public risk for private benefit, and since I've already addressed that, that's all I'll say there.

I guess my bottom line with that is that more thought needs to be given to the whole notion of risk and benefit in the context of naming decisions.

To the second question about design, for examples of peer reviewed research designs that address some of these questions, I refer the audience to the list of references at the end of my presentation which you'll be able to get off the Internet when this is put up there or you could E-mail me and get the same information.

A research design follows from a clear research question and one or more clearly stated hypotheses. Unfortunately these questions and hypotheses have not been clearly stated, or if they

have been stated, there's no consensus on them. Until we can reach consensus on what the questions are and what the hypotheses are that we want to test, we cannot devise rational research designs.

Possible research questions might include:

Does the drug name under consideration present a greater risk of harm due to confusion than the average drug name?

Or does the drug name under consideration present a risk of harm due to confusion that is at or below some acceptable threshold?

Notice that both questions imply a comparison. In the first question, the comparison is to other approved names. In the second question, the comparison is to some threshold.

So one important point is that research designs should include relevant comparators, i.e., controls. This is fundamental to research design and other aspects of the FDA, and yet it's not here in the drug naming. Without controls, there's little

one can conclude from studies of drug names. It's my understanding at present that the FDA and most of the people who talk today, all of the people who talk today don't use any controls.

Imagine if we did clinical trials without controls. What could we conclude? Nothing.

The other measured point I would make about research design is that the design should incorporate state of the art techniques from the relevant scientific disciplines. The most relevant disciplines in the study of drug name confusions is psycholinguistics. Within this discipline, there are standard research designs and measurement techniques for examining errors in visual perception, auditory perception and short-term memory. These techniques should be adapted to to the context of drug names and used as needed.

Sample size. The sample size needed for any experiment depends on the expected effect size of the result and the experimenter's tolerance for false positive and false negative errors in the results.

For example, Flynn, Barker, and Carnahan recently reported in the Journal of American Pharmacists Association that the wrong drug error rate in out-patient pharmacies in the United States was approximately .13 percent, or six wrong drug errors out of 4,481 prescriptions, or 13 out of 10,000, however you want to think about it.

If we were to assume that this were the baseline rate, and if one wanted to detect a doubling in this rate to .26 percent, then assuming a two tailed alpha, it gets technical, but one would need more than 1,570 subjects minimum. The reference I used didn't even deal with event rates this low.

A sample size calculator used on the Internet said you'd need at least 7,620 subjects in each group if you assumed this event rate and an alpha of .05 and you wanted a power of 80 percent.

Belatedly if you wanted to have an 80 percent chance of detecting even one error, if you assume that the error rate is .13 percent, you'd need 1,237 subjects, and the match behind this will be in

my testimony.

From what I've read, the FDA has been using about 100 screeners, and Jerry confirmed this today, and the others use roughly the same numbers, 100 or 200.

Unless the wrong drug error rate is an order of magnitude higher than what Flynn, et al., have observed, then such small samples are unlikely to uncover any errors. If they do, it would be just by chance even if the name is confusing.

Therefore, when it comes to power analysis of naming studies, the very low base rate of name confusions makes realistic experiments difficult and expensive because the required sample sizes would be too large.

In order to do small sample studies, one needs to inflate the error rate artificially by making the task harder or more confusing than it is in real life or also using within-subjects research designs.

In my studies of memory and perception of

drug names, I've always artificially increased the error rate by making the task more difficult either by speeding up the task or blurring the stimuli.

So we can either lose some external validity by artificially boosting the error rate or we can conduct massive studies to detect these very low event rates.

The low base rate for these wrong drug errors shouldn't lead us to believe these are uncommon. Proportionally they're rare, but there's .13 percent of three billion prescriptions; that's 3.9 million wrong drug errors per year. Assuming 60,000 pharmacies, this is one wrong drug error per pharmacy per 5.6 days. So that's what we're dealing with if you can trust these recent estimates.

What should the group look like? The composition of the group of evaluators should ideally be related to the proportional composition of the population of individuals who only count the drug as a professional or a patient. I think the people who spoke this morning have a pretty good handle on this

issue.

Thus the composition will vary depending on the drug's legal status, whether it's a prescription or OTC, its indication, its likely context of use. At a minimum the panel should have a physician, a pharmacist, a nurse, and a patient.

The most meaningful outcome measure is the presence or absence of an error on some realistic behavioral test of memory perception or action.

The next most meaningful outcome is probably an expert judgment on some sort of validated rating scale.

What sort of information should we put in these studies? Studies should include all of the drug attributes that typically are included on drug orders. Again, I think we have a pretty good handle on this. So this is name, strength, dosage form, quantity, and administration schedules. Other attributes that might be relevant but not as critical are colors, shapes, storage, location, outer packaging, indication, pharmacologic category, et

cetera.

The relative importance of these different attributes we have no idea about to be perfectly honest. We have only our intuitions. We have no empirical evidence about which of these factors is more or less important, although Tim Lesar's database may help us get an idea of which of these attributes is most important.

Premarketing and risk management programs. Additional evidence is needed as to the effectiveness of post marketing risk management programs designed to minimize name confusions. Those that have been tried with anecdotal success include labeling changes, shelf shouters, computerized alerts, "Dear Doctor" letters, preprinted prescription pads, and print advertisements. These risk management programs should be evaluated and controlled experiments and real world quasi experiments. The outcomes of the test of risk management interventions should be the difference in error rates with and without the intervention.

Pretest and post test designs are not appropriate here because time itself affects the error rate in unknown ways. In such studies the error rates must be assessed by direct observation, not self-report.

And finally, should OTC and Rx drugs be evaluated differently? I think the answer is no. The issue here is harm reduction. Since both OTC and Rx drugs can cause harm, I think that we ought to evaluate them in the same way.

Thank you for your attention.

(Applause.)

MS. STIFANO: Next we have Beston Jack Abrams, President of ACT, Inc.

MR. ABRAMS: A very happy good morning. I'm very happy to be here because I have the pleasure of representing people who have worked in the drug industry, myself for over 30 years, and are very proud of the contribution I've made to public health.

I'm also happy because I did not do well in college in statistics. I did not do well in

college in a number of other things, but that's beside the point.

(Laughter.)

MR. ABRAMS: The reason I am happy about that inadequacy is that I do not have to deal with some of the numbers and concepts that we've had to deal with this morning and will continue to deal with in the future.

My job, as I find it, is to create trademarks. That's all I do. I do not get involved in evaluating them. That's for others who are more competent in these other fields to do.

But as an agency that develops trademarks, I think we can contribute to drug safety dispensing in two ways.

Number one is to have a deep understanding and an appreciation of what the PhRMA people are attempting to do, and I think we all agree they're attempting to produce safe and effective drugs.

In developing the trademark, if we can

craft it in such a way that the user will have a hint as to what it is intended for, I think we've taken a step forward.

Serendipity has it that this morning two trademarks were presented as potential confusion, one that we would have never anticipated or we would have done something about it, and that is Capoten and Cozaar. And once you see how people, physicians will write their trademarks, we can understand this ensuing confusion.

However, the corollary might be asked how many drug errors were avoided by having a trademark that suggests in some innocuous way the use of it. How many errors were avoided? We will never know. We can only hope that because there is a hint of what Cozaar is inside the name and a hint of where Capoten came from, people will understand what it is and use it properly.

So the question is, yes, errors can be created by improperly employed trademarks or conceivably, they can also be avoided.

The second thing I'd like to suggest that a trademark agency can contribute is the word "perseverance." In this process of developing a trademark, we have heard a legion, a host of processes, hurdles, tests, evaluations, et cetera, et cetera, all designed quit properly to ferret out, to identify a problem before it occurs.

In my experience, roughly 80 percent or more of the trademarks I propose are rejected either for commercial reasons which are quite legitimate or for legal reasons, also legitimate, for safety reasons, et cetera.

The process of evaluating and creating a trademark and working through this process, through the industry, through the FDA and so on is protracted and will test the patience of anyone, but that's the key to what I'm suggesting, and that is when a PhRMA company retains the services of a trademark consultant, it should be understood at the beginning of the relationship that the relationship will not cease until the client has some real assurance that

what they've contracted for is going to be delivered.

The last thing I'd like to suggest is -- and the previous speaker happily touched on it -- and that is the separation of powers, so to speak. I think PhRMA companies should in the spirit of good naming practices should investigate the qualifications of the people that it hires. It obviously will send out site inspectors before they do a clinical research, before a CRO is hired.

And I think the same standards should apply to people who are providing services to PHRMA. What are your qualifications? What is your record? How far do you intend to go with this project? Are you going to finish it? And are you going to do more than you're qualified to do? Are you going to do creative work and are you going to stay the course of the creative process or not?

This sort of thing should be presented up in front so that good communications, good relations between PhRMA and the consultants are preserved, and a better product is the outcome.

Thank you.

(Applause.)

CAPT. PHILLIPS: Thank you very much.

The next speaker is Dr. Suzanne Coffman, Product Manager at NDC Health.

DR. COFFMAN: Thank you.

I'm grateful for the opportunity to address an assembly of look alike/sound alike experts like this.

My job at NDC Health is to provide solutions for retail pharmacies to the clinical issues that face them today, and so I'll be addressing Question 4 on the risk management programs.

NDC Health is a leading provider of health care electronic data interchange and informatics products and services. Two out of three transactions, prescription transactions, in the United States go through our intelligence network, and 90 percent of pharmacies are connected to NDC Health.

If you think there's a lack of information on the incidence of medication errors, that goes doubly or triply for the retail environment. So it's an area that retail is only beginning to look at.

And one person asked the question earlier: how much difference can we make in just looking at drug names and applying the right name to a drug when it's approved?

I don't know the answer to that question, but because there are more than 26 products, I think we'll probably never eliminate the problem completely. So we do need to try other methods as well, and so I'm going to address one of the possible risk management solutions that can be applied after a drug is on the market.

If all pharmacies in the United States used our safety advisor service, we could prevent thousands of just wrong medication dispensing errors in a given month, and that's because we alert the pharmacist during the filling process. We have

already prevented at least 50 errors in the short time that we've been on the market, and also I think this is a key opportunity to use as a risk management tool.

What we do is we send an alert to the pharmacy as they transmit a prescription through our intelligent network. We alert them if the dose that they submit on that prescription is atypical for the drug that is submitted and is also a typical dose for one of the drugs that looks or sounds like that particular drug.

We generate these alerts using a database that we have built on top of the USP-ISMP list of non-look alike/sound alike pairs. We also use the updates from the safety alert newsletter.

We have a U.S. patent pending on our actual decision rules that are used in real time. We recognize -- the reason we took this approach is that we recognize that even though an atypical dose may be within the safe range for a drug, it can be an indicator that there is a problem with a particular

prescription.

We determine our typical doses for each and every drug product, that's every form strength of every drug, and any look alike/sound alike pair by analyzing millions of actual de-identified prescriptions from our data warehouse in Phoenix, and we do age specific typical doses for pediatric patients.

We also have developed a likelihood score actually using the work of Dr. Lambert. We look at the Levenstein distance for similarity of drug names. We look at the comparative frequency with which the drug is prescribed, which we use as an indicator of how comparatively familiar the pharmacist is with the two products, again, at the individual dosage strength level, and then we also use whether there are any same strength or look alike or sound alike strengths available between two given drug products to tell us the likelihood of a look alike/sound alike error.

And the service is easy to turn on and

off. We do it at NBC Health as soon as the store agrees. So there's no implementation on the part of the store. There's no hardware or software to install.

As I said, we have already prevented 50 potentially clinically significant medication errors, and we have only been up and running since really January 1st, and that's in fewer than 200 stores so far, though I'm hoping for many more by the end of the year.

We are actually -- let me give you some errors we have prevented. We had a changed from Claritin D 12-hour to Claritin D 24-hour; isosorbide dinitrate to propranlol. That's the isordilandral (phonetic) pair.

Two changes from same strength of hydralazine to hydroxyzine in two different stores. A change from Lamictal to Lamisil. A change from glyburide to glipizide.

And then among the non-look alike/sound alike drug changes, we've had a number of changes,

say, from a Ditropan XL to an immediate release version of the same drug.

We've had changes in strength of Avandia, lisinopril, prednisone.

And then we had a very interesting drug changes that's not a known look alike/sound alike pair. We had a change from Elavil 25 milligram to Ativan .5 milligram with the same quantity end date supply.

And we also had a change of atenolol from half a tablet a day to two tablets a day.

So that's just a few of the 50 that we've prevented so far, and again, that's since January.

We are doing a controlled study of the impact of our service in a 115-store chain, and the way that we are determining whether there are any changes made is by taking the initial transaction that generated the alert in the first place and then the immediate subsequent transaction and determining whether there was any change in the drug quantity or day supply as a result of the pharmacist receiving

our alert and double checking the prescription.

I also think, as I mentioned that this tool could be potentially useful as a risk management tool in post marketing surveillance. It would be useful in testing the findings of the name screening process premarketing.

We can send alerts either on all prescriptions for that new drug for a very short period of time or only one that doses atypical.

We can quantify the results because we collect all of the data both when alerts are generated and also when there is a potential problem, but an alert was not generated.

We can report that data. We can track it. We can categorize it by day of the week, by store prescription dispensing volume, any which way that would be useful.

And I think it might potentially be an alternative to last minute name changes. And I can tell you that I am a pharmacist, and I have made look alike/sound alike dispensing errors, and I would love

for a system to send me an alert telling me that I might be in the process of making one.

To just to summarize again, I think it's a very useful tool. If it were in use in all pharmacies, we could prevent thousands of medication errors every month. We send alerts during the filling process. We've already prevented at least 50 clinically significant errors, and I think it's potentially a valuable post marketing surveillance tool.

Thank you.

(Applause.)

MS. STIFANO: Next we have Kasey Thompson, the Director of Patient Safety with American Society of Health Systems Pharmacists.

DR. THOMPSON: Good morning. My name is Kasey Thompson. I am Director of Patient Safety of the American Society of Health Systems Pharmacists. ASHP is a 30,000 member professional association that represents pharmacists and scientists that practice in hospitals and other components of integrated

health systems.

We are grateful to the FDA for calling this public workshop to receive input on the agency's approach to minimizing medication errors through improving the drug naming process.

Section 3(f) of the FDA's recent concept paper entitled "Premarketing Risk Assessment" discusses how drug sponsors can minimize medication errors. Specifically the station states ideally a sponsor would conduct a risk assessment to insure that a product's proprietary name, established name, container label, carton labeling, package insert, and/or packaging do not inadvertently contribute to medication errors.

For example, a sponsor could perform a medication error prevention analysis to minimize the potential for an error through corrective action, including renaming, relabeling or repackaging.

The concept paper goes on to state that sponsors should assess a product's naming, labeling and packaging by obtaining first hand information

from physicians, pharmacists, nurses, and consumers. This sponsor initiated assessment would help to minimize medication errors and help speed FDA's review of these issues.

At a public meeting on risk assessment last April, ASHP strongly supported inclusion of this language in any further guidance document related to premarket risk assessment issued by FDA, and we urged the agency to quickly implement this concept.

We have been encouraging FDA to do this for a very long time. In September 1998, we stated at an FDA professional organization meeting that drug naming, packaging, and labeling was a critical issue that had not been adequately addressed by the FDA despite the fact that there had been abundant evidence that poor product design is a major contributing factor in medication errors.

At a meeting in February 1999, we stated that one solution to the problem of medication errors stemming from poor package design and nomenclature is to require real life submissions from the

pharmaceutical industry prior to drug approval, and that before the FDA approves any new drug or biological product, it should require manufacturers to document that it has rigorously tested all packaging and labeling before naming for their potential to induce errors and patient harm.

This testing should be done using proving methods involving practicing pharmacists, physicians, and nurses in simulated work environments.

In May 1999, we commented that the FDA has an obligation to quickly review and revise its procedures to eliminate medication errors that occur due to look alike and sound alike names, similarities in packaging and other labeling and packaging problems.

We also noted that patients should be considered the partners of health professionals in eliminating medication errors, and they should be involved in providing input into the safety design of drug product labeling.

We are pleased that the FDA concept paper

includes a provision for patient-consumer input.

In January 2002, in comments to the agency on its performance goals for the reauthorization of the Prescription Drug Marketing Act, we stated that the most consistent message ASHP hears from its members is that the FDA should be doing more to insure that drugs are safe for patients and that safety issues must be anticipated through premarket evaluation.

One specific new performance goal that we recommended was for the FDA to engage pharmacists, physicians, nurses and human factors experts in documented failure mode and effects analysis of prospective product nomenclature and labeling to minimize the opportunities for sound alike names and look alike packaging for causing medication errors.

In terms of the specific questions that the FDA asked participants to address for this public meeting, ASHP has the following comments.

Question 1, are methods currently employed by sponsors in FDA appropriate for

evaluating look alike and sound alike names?

Generally the kinds of methods being used by the FDA could detect naming problems. Our concern is to what extent FDA staff stimulates the range of real life drug order situations common in hospitals and health systems.

Mobility brings together physicians and pharmacists from different regions of the U.S. with characteristic dialects and from other parts of the world with primarily languages other than English. Face-to-face and telephone communications are easily confused by these differences.

The methods and forms of medication order writing capture and transmission vary considerably among hospitals. Orders can be handwritten, imbedded within progress notes or segregated on distinct order sheets that separate the drug name from indication.

Orders are transmitted to the pharmacy by NCR copies, internal fax machines which confound handwriting variations with smears and electronic artifacts, and a Susan Winckler mentioned, confusion

in errors caused by electronic order systems should also be considered. This is very important.

And let us not forget that hospital and health system patient populations are also becoming more culturally and linguistically diverse. Communications with patients about their medications is an important component of medication error prevention.

Question 2, which deals with how studies are designed to evaluate potential prescription errors. Study design should, to the extent possible, replicate common medication order situations with experimentally known vulnerabilities for error. Designs should include multiple detection and interception methods as appropriate for the vulnerabilities in each step of the medication use process.

Expert committees should be representative of those health professionals, especially physicians, nurses, and pharmacists who have essential roles in hospital and health system

medication use processes.

Question 3, what kind of information should be included in verbal or handwritten prescription drug studies?

Information requirements alone are insufficient. How medication orders are communicated and the context in which they are communicated either contribute to or reduce the potential for errors. Studies should look at error potential of proprietary names alone in the context of typical medication orders and standardized medication orders that incorporate requirements known to reduce a likelihood of misinterpretation.

Question 5, should there be different trade name evaluation procedures for different classes of drugs?

There is no difference between prescription and non-prescription products as far as error potential for interchangeability and subsequent patient harm. ASHP would like to emphasize the importance of name recognition for high alert drugs,

such as anti-neoplastics and other hazardous drugs that have very low therapeutic indexes and, therefore, a high probability for patient harm if an error occurs due to name confusion.

ASHP believes that the FDA is taking the right approach to this serious public health issue and appreciates this opportunity to present its comments relating to the FDA's program for minimizing medication errors.

Thank you.

(Applause.)

CAPT. PHILLIPS: Thank you.

Our last speaker is David Wood, CEO of Interbrand Wood.

MR. WOOD: Thank you for allowing me to come here today to spend a few minutes with you. I'm going to make some comments. I have nothing written down, no notes. You won't find anything on my Web site. So you're going to have to pay attention.

(Laughter.)

MR. WOOD: When we first started to get

into the business of evaluating and testing and assessing new names for drugs in the mid-'80s, it was beyond our wildest expectation that 15 or 18 years later we'd be here in Washington with two or 300 people, with the highest levels of the FDA, discussing things that we didn't even know existed in those days like errors and misprescribing and all of that kind of stuff.

If Bruce's statistic of .13 percent is correct, we have succeeded beyond our greatest expectations. We have done an extraordinary job in a very, very difficult circumstance in reducing error to apparently an almost negligible percent in an environment which is designed specifically to cause error.

The prescribing chain in our industry is extraordinarily difficult, and I would suggest that we will never ever reach zero error.

Timothy Lesar said something this morning that resonated strongly with me, and he said this. Risk assessment must include multiple drug

characteristics, not just names.

I would add two words to that. Must include multiple drug and brand characteristics, not just brand names.

Don't let's make names the whipping boy for an industry and a system which needs to pay attention to many things other than simply the brand name. We must pay attention to every aspect of the prescribing process because if we don't, we will simply squeeze all of the juice out of brand names and allow the other components, which are contributing to error to continue to contribute to error.

We must allow the sponsors in the pharmaceutical industry some latitude to name its products and build its brands, obviously cognizant continuously of the need to protect the public, but they must be allowed to build their businesses and build their brands within reasonable constrictions, not unreasonable constrictions.

Risk assessment is exactly that. It's

risk assessment. We'll never get to zero. We must assess risk and then make mature, adult, informed opinions based upon that assessment. That's all it is.

So zero risk doesn't exist. We won't achieve it. Let's be reasonable in how we assess the risk.

Multiple drug and brand characteristics. I would challenge the industry to do a far better job of building its brands. We have the brands are the ultimate global shorthand. If I stand here and say to you Colgate or Colgatte (phonetic) or Tylenol, pictures come into your mind whether you use those products or not. You know what they are, what they do, what they look like, et cetera, et cetera.

Unfortunately, we as an industry are not terribly good yet at building brands, and it's relatively unusual that we can say a brand name and a picture of what that brand is comes up for us. If I say Viagra, we get triangular and blue. If I say Lipitor, I'm not sure what we get.

We don't know about packaging because most of our products are not packaged. We are probably one of the few industries on earth that negates or dismisses the opportunity to package its products at the point of delivery. So we allow our product to go out in the sort of generic form of big containers and so on, and the point of delivery, we allow it to be put into no name, no personality, all the same vials, et cetera.

So we have a lot of components of our business which we need to pay attention to in order to impact what we're talking about today, which is reduction of error, safety for no look alike/sound alike, et cetera.

Look alike can be look alike in many ways, not simply the name.

I'd like to congratulate everybody for being here today and ask one question. What took all of us so long? We should have done this ten years ago, 12 years ago, 15 years ago. What took us so long.

For those of you here from Europe, I would strongly encourage you to get hold of EMEA and do this sooner rather than later because everybody will benefit from it, and I don't want to prior to lunch be standing up here sort of berating things, but I just wanted to have an opportunity to perhaps put a little different spin on something and not, as I say, make names the total whipping boy.

I have no other things to say other than it's a pleasure to be here, and I thank you for inviting me, and I look forward to seeing you all at lunchtime.

(Applause.)

CAPT. PHILLIPS: Thank you very much for a stimulating discussion, and it's lunchtime. We're going to adjourn for one hour. So we should be back here at ten minutes to one, and we have three restaurants in the hotel and there's a food court also. So lunch is on your own.

And we look forward to seeing you back at ten to one.

Thanks.

(Whereupon, at 11:48 a.m., the meeting was recessed for lunch, to reconvene at 12:50 p.m., the same day.)

AFTERNOON SESSION

(12:50 p.m.)

DR. COHEN: Good afternoon. In the interest of keeping on time, I'm going to start this next section. So thank you for returning to your seats.

There are two other people that we work with that weren't mentioned this morning who helped to put this program together. Ms. Sharon Olmstead from Pharmacia corporation worked on behalf of PhRMA in planning this meeting, had quite a bit of input, and probably the person that really helped to pull all of this together is a fellow named George D. Di Domizio, whom most of you probably know. George has been instrumental in the whole area of trademark medication safety with trademarks, and he, too, participated in putting this meeting together.

The purpose of this afternoon discussion is to begin to explore the methods themselves that you heard about. How can they be improved?

I don't think anyone would argue with the

fact that we would love to see a better way of doing this process or improving what we're already doing. No one would argue with that.

And so it's my pleasure to introduce our first independent expert panel.

Dr. Brian Strom is from the University of Pennsylvania School of Medicine, and he's going to discuss issues surrounding this sampling issue in order to screen proprietary drug names.

And Ms. Shari Diamond is from Northwestern University School of Law, and she's going to discuss the pros and cons of using questionnaires as a screening tool. You heard that that was done by all except FDA.

And next we have Kaz Jaszczak, and he's from Parascript, LLC, and he will be discussing handwriting and voice recognition models.

There will be a short time for questions and answers after these formal presentations, and I would like to remind the speakers that we are going to hold you to the time frame allotted for your

remarks because it's very important that everyone be given a chance to do their presentations in the time they were given.

Our first speaker will be Dr. Brian Strom, and again, this is on sampling issues, and we have a couple of questions that we posed to him to cover.

What is an appropriate sample size of respondents to best determine the risk of sound and look alike proprietary names in the prescription drug study group or in a focus group or in a survey group?

There are some organizations that do this type of testing that would use, as you heard, in the hundreds of respondents, and others use a much lower number. What is an appropriate sample size?

And then should the sample be randomly selected? Is it important to have a statistical significance for this type of evaluation?

Dr. Strom.

DR. STROM: Thanks, Mike. And I'm not sure I thank you for the position on the program,

being right after lunch. Not only does everyone wander in late, but they're also half asleep, and trying to present a topic which is theoretically statistical on top of that in that time slot is particularly challenging. But I'll certainly try to keep you up.

Again, my name is Brian Strom. I am Chair of the Department of Biostatistics and Epidemiology, but I'm not a biostatistician. I'm an epidemiologist.

I also am new to this field of drug names. I'm a pharmacoepidemiologist. I study drugs and adverse reactions and certainly study patient safety. I'm also principal investigator, along with Bill Campbell of one of the CERTS, the Centers for Education and Research in Therapeutics. And I'm also on the Drug Safety and Risk Management Advisory Committee of FDA.

So I have a lot of interests that surround this, but have never been involved in this issue that perhaps can be useful in terms of bringing

the perspective of an outsider who understands something about research methods, but doesn't know anything about how to choose drug names.

As Mike said, the two questions I was asked to address was what is an appropriate sample size of respondents to best determine the risk of sound and look alike proprietary names in the prescription drug study group and a focus group in a survey document.

And, two, should the sample be randomly selected? It's important to have statistical significance for this type of evaluation. In many ways, another way to look at this is shown here. If people can't read it, it's as well. It began, "Well, I'll be damned if I'll defend to the death your right to say something that's statistically incorrect."

It's really a question of how do you apply statistical methods to the kind of questions we're addressing today.

I would like to begin by thanking Sean Hennessy from our group. I am a complete outsider to

this area, as I mentioned. Sean is a pharmacist and epidemiologist who has been at least a little bit more involved, and a lot of the ideas I'll present to you today came out of a couple hour brainstorming session that Sean and I had together.

What I'll be talking about is first a brief introduction, then a very brief discussion of very general principles of sample size calculations and sampling the two specific questions I was asked.

And for those of you in the audience who are researchers, I apologize for the simplicity of that, but I assumed correctly, it seems, that many of the people who would be coming wouldn't necessarily be researchers.

I will be talking about applying those general principles to this situation and spend most of my time making a series of recommendations for research to guide the future.

Well, the basic designs used in study designs in general are shown here. There's what are referred to in analytic studies and what are referred

to as descriptive studies. The analytic studies include the experimental trials, the randomized clinical trial that this group is undoubtedly most familiar with.

Also, cohort studies and case control studies. Descriptive studies include an analysis of secular trends, case series, and case reports, case reports being analogous to the MedWatch type of spontaneous reports.

Just to be sure everyone is comfortable with the distinction, cohort versus case control studies, both cohort and case control studies are intended to give the same basic information inherent in this two-by-two table, that is, whether an exposure is present or absent and whether a disease is present or absent.

The difference is a cohort study approaches it horizontally recruiting people in the study on the basis of presence or absence of exposure, and then the process of the study looks to see where there's a difference in outcome.

In contrast, a case control study approaches this vertically, recruiting people into the study on the basis of the presence or absence of disease, that is, are there cases or are they controls?

And then the process of the study is looking at any differences in antecedent exposure.

Both of these approaches though, cohort and case control randomized trial, which really in many ways is a subset of a cohort study. You're just randomly assigning people between the two groups. All require the use of a control group, and I'll come back to that in a minute.

Well, in this context, how do you calculate a sample size? How do you calculate what you need?

And basically whether you're talking about a cohort study or a case control study there are basically five related variables, and if you calculate -- if you're given one of them you can calculate or if you're given four you can calculate

the fifth.

One of them is alpha. What is the probability -- the conventional of this is .05 -- your willingness to accept a false positive study?

One of this is beta, which is conventionally .1 or .2, which is your willingness to accept a false negative study that is missing a real difference when a difference really is there, and again, talking about difference between the exposed and the control group or between the case group and the control group in a case control study.

One of them is a measure of variability or precision in the measure, commonly standard deviation if you're dealing with a continuous variable, and the last is the delta or how small a difference do you want to be able to detect.

And the smaller the difference you want to be able to detect, the larger the sample size that you need.

These five variables, you basically specify any four of them and you can calculate the

fifth. So the answer to how big should the sample size be is if you give me those four numbers I'll plug it into a formula and give you the answer. But you have to specify those four numbers.

In a cohort study it's analogous, alpha, beta. The two additional analogous variables are the incidence in the unexposed control group, that is, how often normally does this disease occur in the unexposed controlled group, and then the delta becomes how small a relative risk do you want to be able to detect.

And in a case control study that's here, again, alpha, beta, and you're looking at the prevalence of the disease in the undiseased control group -- sorry -- the prevalence of the exposure in the undiseased control group and the delta is how small a difference do you want to be able to detect.

So, again, in principle, sample size calculations are simple, their mathematical formula. If you specify the variables, you can calculate it accordingly. The key issue is specification to

variables, and I'll come back to that.

In terms of the question of sampling, the overall approach -- there's no pointer here, I guess. I'll try to talk through it -- but the overall approach to a study design is shown here. You choose a study sample that is theoretically a random sample of a general population. The generalization from that study sample to that general population gives you, if you have a statistically significant finding, you have an association.

That study sample, again, in theory is a random sample. In practice, it virtually never is, but people make believe it is and do the analyses, making believe it is a random sample.

The second step, and to help you in that steps you have all of the biostatistics and all of the rules and regulations and formula related to biostatistics.

The second step is more subjective and that is biologic inference, going from a statistically significant finding in a given study to

a general conclusion about scientific theory or causation. That's more subjective. You don't have anything as precise as biostatistics to inform that step.

So you might have, for example, a study of middle aged white men, all of whom have high blood pressure and you randomly assign half of them to get methyl dopa, to choose an old drug that many of the companies probably are part of, and half of them to get placebo, and you look to see what happens to the blood pressure in the two groups.

In the methyl dopa group the blood pressure will go down. In the treatment group the blood pressure will also probably go down due to regression to the man, though probably not as much. And if the difference between those two groups is larger tha you'd expect just by chance, that is, you have a P value of less than .05, you have a statistically significant finding. You have an association. The conclusion is methyl dopa lowers blood pressure in middle aged white men.

To then take it the next step and say methyl dopa is an anti-hypertensive drug is a totally separate subjective judgment. You're generalizing to women. You're generalizing to the elderly. You're generalizing to young. You're generalizing to other races. You're generalizing to all sorts of groups that aren't represented, and that type of generalization is more subjective in judgment.

There are a set of criteria to assist that kind of judgment. Actually a variation of them was first put forth by R.A. Fisher -- no, sorry -- by Sr. Austin Bradford Hill in the late 1940s. Probably the best known description is in the first Surgeon General's report on cigarette smoking and cancer.

But the bottom line is that's subjective. The key thing we're talking about here is the top part, which is statistical inference. Key to the question of statistical inference is is the difference between the two groups larger than you'd expect by chance.

Well, let's take these general principles

to the situation here. The central principle of research design which my trainees get very, very tired of hearing me say because I say it all the time is the question is what is the question. And the issue here in this situation is that there are no a priori hypotheses being tested to be able to consider sample size calculations or questions of sampling.

What I heard described, talking to Mike beforehand, talking to Sean beforehand, listening this morning, is essentially qualitative research. Some of it is quantitative in what's being done, but there's no comparison. There's no exposed group and control group. There's no disease group or undiseased group. There's no a priori hypothesis. There's good reason people are left with, well, should the sample size be 30 or should it be 100 or should it be 1,000 because the answer depends on the question that's being asked, and I'm not hearing a specific, definable question in any given situation.

One of the things that, again, I harp on with my trainees is if there is a question about

study design, the answer isn't to focus on the study design. The answer is to focus on the question, on the scientific question.

And if you refine your question enough, the study design answers become very easy. And so how do you calculate a sample size with those four variables I gave you, given the kind of efforts and questions we heard about this morning? It's not even a meaningful question.

And so what I would suggest is that the key thing is that we evaluate the current process in a quantitative way, and there's a number of aspects of that.

Part of what was striking listening this morning was the striking lack of consistency in methods used by different people with absolutely no evaluation that I heard about which one was right, and so there's no way to answer what should the sample size be and whether it should be sampled or any other research design question if you don't know which one is right.

And so what I would argue is you need a four step process, and I'll talk about each of them very briefly.

First is to standardize the procedure.

Second is to test for reliability or reproducibility.

Third is to test for validity.

And fourth is to make changes in the procedure accordingly.

Firstly, standardize the procedure. We heard a host of different approaches this morning. One needs to choose among the current possible approaches a standard to be evaluated more rigorously.

You can choose more than one standard. You can choose any. You know, that's not the point. The point is an evaluation of one of the approaches is generalizable to that approach only. It means that approach does or doesn't work or is or isn't reliable and so on. It doesn't tell you whether any of the others do.

And so what I'm going to describe to you has to be done one approach at a time to see which approach works, if any of them, and even for that approach it has to be standardized. In order to be able to evaluate something scientifically, you need to be able to know what it is. When you're evaluating the efficacy and safety of a drug, you need to know what's in the pill.

So in order to be able to evaluate whether or not this process of evaluating drug names works, we need a very precise description and specification of what that process is, and then an evaluation of it accordingly the way I'll describe, but a different set of evaluations than would be necessary for each different permutation in the way that this process is handled.

Once you have a process that is the standard process, that is, a standardization of one of the processes that you want to evaluate, the first step to look at is reliability, and what reliability really means is reproducibility. Does it give you

the same answer? If you do it twice do you get the same answer? If you do it 100 times, how often do you get the same answer?

Evaluate the same drug names in the same process with multiple different groups of survey prescribers and different groups of experts in order to see whether there's adequate agreement.

Certainly we heard a lot of suggestions this morning about lack of agreement, lack of agreement with the outside firms versus FDA, you know, lack of reliability in multiple different ways.

And indeed, if there's no reliability, validity is impossible, and the procedure should be abandoned. There's no reason to be running an exercise if it's not reproducible because what's the purpose if it? You can't get the same answer over and over again.

And what that argues, if there's no reliability, no reproduci/bility is to go back and try to change your standard and make it more precise and change it in a way that will allow it to be

reliable, and there's no reason to go any further in this evaluation process if it's not reproducible because all you're doing is analyzing noise, is analyzing random error.

Validity, to test for validity, you need a gold standard. What validity is it's saying how does this measure compare with the gold standard.

And the question in this field is what would you consider as possible gold standards. Well, I'll throw out briefly three different possibilities. One are drug names that were rejected in the initial FDA review.

Second are drug names that were withdrawn due to problems once a drug was in the market.

And third is a direct measurement of the error rate.

So one possibility is drug names rejected in the initial FDA review. That clearly is, if not a gold standard, maybe a silver standard. To the degree companies are looking to try to second guess what the FDA will do, that can represent a standard.

But the problem is was the FDA's original decision correct. You don't know that, particularly given the whole process is as subjective as it is.

(Pause in proceedings due to electrical failure.)

DR. STROM: So the issue of the FDA decision being a gold standard isn't a gold standard for what matters to patients, but at least from a regulatory and from a commercial point of view, you could try other measures, these other approaches versus an FDA decision using a history of FDA decisions in the past as a perhaps silver standard.

The second possibility are drugs names that are withdrawn have been withdrawn after being marketed due to problems that occurred. The potential issue there is that the knowledge of the reviewers could be problematic. That is, the reviewers themselves could well know that these drug names are withdrawn, and so your process would be, of evaluating the drug names, would be flawed because they would know that these are problems.

One approach with this is to use data from other countries, drug names that were withdrawn, never marketed in the U.S. and withdrawn in other countries rather than here.

Another would be to use drug names that were withdrawn here years ago and use on your panel young pharmacists who might not know that history as a way to get at that.

Of course, you're still left with a question of was that withdrawal decision a correct one.

The most direct way and the way I would argue makes the most sense is to try to have a direct measurement of error rate, to simulate a real life situation in this study setting.

Again, this would need to be done differently. What I'm describing here is for looking at written prescribing. Verbal orders would need to be evaluated differently. Any other approach would need to be evaluated differently. Again, the question is what is the question, and you need a

different set of evaluations for each.

But what you could do is you could choose good and bad options for new names; enter those possible new names into the standard prescription entry order program; ask large numbers of physicians, ideally randomly selected, to write orders for those drugs; ask large numbers of pharmacists, again, ideally randomly selected, to fill each script by entering that script into their prescription order system, and directly measure the resulting error rate.

How many errors are actually made when you simulate that situation? Measure it directly.

This would not be an inexpensive study, but it wouldn't be outrageously expensive. It could potentially be the standard approach you ultimately use, though it would be nicer to be able to not have to do that with each and every drug and each and every drug name, but rather just use this as a gold standard and use that as the gold standard to evaluate the other approaches by.

Then you change your procedure accordingly. Determine the appropriate cut point for expert ratings by doing an ROC curve for gold standard. What we heard this morning is there isn't even standardization about what kind of ratings people give, whether they're presented the ratings. Certainly to the degree there's ratings, how bad a rating is too bad? What is the rating that, in fat, does predict medication errors?

That basic kind of information isn't there and could easily be derived from this kind of study.

You also could determine the appropriate sample size through simulation. That is, how many people need to be in your focus group in order to achieve results consistent with that gold standard? Because then you could basically run your expert evaluation with five experts, 15 experts, 30 experts, 50 experts, 100 experts, and now you have a gold standard, and looked to see how many experts you need in order to reliably give the correct answer and then

modify the processes accordingly.

Again, this isn't an inexpensive effort. It could potentially be funded using FDA extramural funds, using AHRQ patient safety funds, using National Institute on Aging as a pharmacology program. NIGMS has a pharmacology program. From what I heard this morning, there might be interest in PhRMA in developing a better approach. The testing companies; some combination of matching. There's lots of ways this could be done, but to a real degree the field hasn't started at square one in terms of being able to answer the sample size questions.

So my conclusions are that applying a quantitative approach to evaluating what has so far been a qualitative one could lead to major changes in the procedure and major improvement in the net results.

The alternative is here: to my data, right or wrong.

Thank you.

(Applause.)

DR. COHEN: Thank you.

Our next speaker is Shari Diamond, and these are the questions that we posed to Shari.

When constructing a survey form to determine trademark safety issues, should questions be multiple choice or open ended?

Should the questionnaire be self-administered by respondents?

Are there situations where focus groups are preferred over individual respondents to evaluate new drug names?

How much information should a respondent have about the trademark being evaluated?

A question for this purpose: how do you insure the reliability and the relevance of the data being collected?

Shari.

MS. DIAMOND: Well, since Brian started out with confession time, here's my confession. My confession is that I'm something of an outsider here as well. I testified in a trademark case as an

expert because I had done a survey in the case, and that's where I've learned most of what I've learned about the pharmaceutical industry, except working in my Doctor Dad's office growing up. So that's where I come from.

I do teach intellectual property and did practice trademark law for a little while, and I have a Ph.D. in psychology, social psychology. So I like to think I know something about research design.

And so I was trying in preparing for this set of questions that I was given to find out what has happened in the rest of the industry and what was going on and have pieced together bits and pieces of information about it, but it is clear to me that Brian is correct and from this morning, that there's a lot of variation out there.

So what I did is I'm going to imbed the questions I was asked to address in a little bit broader topic, and I called it research design and questionnaire structure, which I'll go through, but it allowed me to talk a little bit about control

groups as well as other things, but I'll skip that since Brian talked about that as well.

First of all, I want to say that the challenges to designing a test for these problems are huge. When I got involved in looking at these problems of medical error, prescription error, I became aware of just how difficult it is for products that are not yet on the market, in particular.

Some of what Brian talked about about having people fill prescriptions, well, there's no prescription to fill. So simulating things is pretty hard, and you have to simulate a variety of different things. You have to simulate written prescriptions for drugs that don't yet exist and so, therefore, having doctors fill out prescriptions for drugs that don't exist, well, they have to simulate how they would fill out a prescription for that drug.

The same thing for things delivered orally for a drug that is not yet being marketed, and the same thing with regard to filling by pharmacists. So it is a terrifically difficult thing to study,

and I'm glad that the FDA and others of you have taken this on because, of course, the National Academy of Science's panel on error rates was a little daunting.

Well, I took a look at what the FDA had been doing, and the notion of the expert panels that begin all of these are really a good place to start. I think compared to Brian I may be more sympathetic to some of the expert panels as a source of information because the expert panels are, in fact, knowledgeable about currently marketed drugs in a way that probably nobody else is knowledgeable.

Similarly, they're familiar with the drug pairs that have generated errors in the past and also, as we heard this morning, there is a lot of use of source lists to generate potential candidates with confusing name similarity, and that is all accessible and familiar to the expert panels.

And Robert Sternberger, who is a psychologist, has written about the issue of tacit knowledge, so that even if expert cannot specify

exactly how they know things, can't really explain it, they do have frequently a fund of knowledge that's very useful in making judgments.

On the other hand, there are limits on the ability of experts to predict errors. After all, it's not the experts who are making the errors. It's going to be the other folks out there who are doing the prescribing and filling the prescriptions, in particular, filling the prescriptions.

And they may very well generate many similars that don't really pose a threat in the ordinary situations in which people fill prescriptions, and they also may miss potential errors. And I just give a couple of examples of situations where that might arise because the experts don't generate mispronunciations that actually occur in the field and cause error, or they may not anticipate similarities generated by handwriting.

So those are two possible ways. There are, of course, others. We know that people are not always very good judges of what causes their own

behavior, let alone what causes other people's behavior.

So we need to test expert predictions, and the testing phase for gauging actual reaction is crucial. Now, you need a sample drawn from a relevant population, and Brian talked about sampling. So I won't talk about that. It is a daunting thought for a large enough sample to collect very low base rate errors, which is what presumably occurs in this situation even when there is a medical error problem.

And they have to be responding to appropriate stimuli, and the third speaker on our panel is going to be talking about the handwriting. So I won't talk about that.

But the design, assuming you want to test the name, Taxol, as you know, Taxol had a confusion problem or at least it was potentially a source of error with another chemotherapy drug.

So we would set up a situation in which respondents are told that they'll see a series of

drug names one at a time, and they have to be told that some drugs are currently on the market and some are not yet available so that they are prepared for the possibility that there will be something there that they don't recognize in terms of the testing procedure.

And so explicitly procedures for testing. They might get a set of handwritten drug names one at a time, right? So this can be done by self-administration, one of my questions. If respondents are hooked up to the Internet, as most pharmacists are, so it's quite possible to be able to test this.

And this actually looks a little bit like some of the programs that are in place that we heard about indirectly this morning.

Now, this one is a little trickier, and we really need some research on this because one of the things you like to do is have a timed exposure to reflect the usual time spent in examining a prescriptions. If we don't have that research up to this point, we need that research because one of the

concerns that I raised earlier was about the prospect of having simulated circumstances that don't accurately reflect what goes on when prescriptions are actually being filled.

And if people are doing a test and they know that they have as long as possible to do it, they may do it more carefully than they would be able to do it in the rush of activities in the ordinary pharmacy.

So the wonderful thing about a computer is obviously you can limit the time, and similarly, the order of presentation of a series of names can be rotated so that it isn't the first one always. It isn't the second one. It's a series, and you can balance for order, a very good piece when you're doing this kind of research.

The names that are shown apart from the critical names issue, a kind of control for the ability of that pharmacist to recognize various drug names in that procedure. So here's an actual instruction. You'll be shown the number or the names

of several drugs one at a time. Some of the drugs may be currently on the market and some may not be.

For each drug, the name of the drug will be followed by several questions, and these questions will ask your reactions to the drug name you just saw. Just pretty straightforward stuff.

And the questions after each name is shown, please type in the name of the drug you just saw, or if it were administered by an interviewer, what is the name of the drug you just saw? Could you spell that for me?

And then a few follow-up questions. Now, there can be many more follow-up questions than the ones that I've identified, but I've identified a couple that I thought would be useful for tracing sources of difficulty, and those were:

Have you seen this name before today? Because people who think that they have seen the name of the drug before today when it isn't already available are engaged in a minimal form of recognition, and the follow-up to this:

If yes, do you happen to recall what conditions it is used to treat? So that is the follow-up for finding out. Obviously more questions can be added to this kind of a protocol.

Now, there was a question about other cues that might be used, whether it's just the name that you want to test or whether you want to test other kinds of information that might appear on a prescription.

Obviously testing the name alone maximizes the likelihood of name confusion. If you're using relatively small samples, you may want to do that to detect low levels of error rates.

There are other cues, of course, that you can use which should reduce apparent confusion, and in fact, the best prevention of error is to provide multiple cues, and I didn't hear anything about this this morning, but one of the things, again, in my outsider capacity in the literature I read up to this point, I was dismayed to see that there really are a number of potential remedies or prevention techniques

that should be more widely used.

We had some discussion at lunch, and George mentioned that he carried a prescription, I think for his wife if I'm correct, and he noticed on the form that it had the doctor's signature line, and it had, "Please print your name below." Right?

Of course, it didn't say that for the name of the drug that was on the prescription. So that was scrawled in an almost illegible form, but the doctors name was nicely printed below the signature.

Engineers understand this. They build in normal redundancy and cross-checking, and there are a variety of methods that have been suggested in the pharmaceutical context for this, like having the generic name as well as the brand name on a prescription or indicating the way in which the drug would be used, but we haven't gotten to that point so far.

So including cues in the screening test may reduce apparent likelihood of error, but it won't

reflect reality if cues are inconsistently provided, which apparently they seem to be.

So I would lean back on paying attention to the name itself until we can get to the point where multiple cues are dependably introduced.

Another approach that might be taken to the design approach I mentioned to you earlier. When there is an expert panel identifies a particular similarly named drug, this is what I would call a line-up procedure, and we see it. It is really like a line-up, and we sometimes do this in the trademark area in assessing confusion, and you show the line-up of products after the person has seen and, again, tell them that the drug may or may not be displayed here and indicate whether it's in the display, and if it is, which number is it.

This is particularly likely to get chance identifications, and so, therefore, it's crucial to control for guessing, to introduce a control group methodology. And you might want to confine this procedure only to situations where similarly named

drugs are likely to be stored side by side. So that would be most likely for that situation.

Well, another one of my questions was: should the questions be close ended? This is a little like sampling size.

It depends is the answer to the question. Right? It depends on what you're asking. If you're trying to see if a person can reproduce the name of the drug that you have shown to him or her, you don't want to provide a multiple choice set of possibilities for him or her to choose among because that is loading the dice, making it easier to identify.

But if on that third question I mentioned you had a list of conditions, and you were asking somebody, well, what condition would it be prescribed for, you might supply a whole list, a fairly comprehensive list of conditions and ask them to check off all where it would be applied, and that is a close ended question.

Line-up is essentially a multiple choice

question. It's a recognition task. You're asking which of these, if any, and in that context, a close ended question makes sense, but you have to be sure to have a control group introduced.

We haven't talked in the course of the day so far about focus groups, but my questions included a reference to focus groups, and so focus groups presumably were being referred to as a substitute for the testing that I've just described.

Focus groups, in general, are good for generating idea, and the expert panel is exactly like that kind of focus group, of generating ideas, of feeding off one another as they talk about things that might be a problem, and generating a series of possibilities.

But they are weak for evaluating individual reactions to specific stimuli. Part of the problem is the interdependence of the responses from the group members because the joy of it is that they influence one another.

And the second part commensurate with

that is that that means that unless you're going to do a large number of focus groups, you are going to have a very small n. That is, the group is really the unit because of the interdependency, and of course, there's a crucial role that the moderator plays in terms of influencing the structure of the focus group and has to be carefully monitored, as well.

Recently a district court judge in one of the federal district courts took out after a company not in the pharmaceutical industry, but another company who produced, quote, survey results that consisted of focus groups.

Problems in validating. This is something that Brian touched on, and if you look at the reality of what's out there and the testing that currently takes place, there's a kind of one sided partial and incomplete feedback. So we have our approvals followed or not followed by reported medical errors, and so we know that if something was permitted to go on the market, we get a feedback on

whether it, in fact, produced a number of medical errors that were reported.

The problem is it's a very one sided testing mechanism, in addition to being incomplete and depending on reporting, because we don't know anything about the disapprovals because they were never out in the marketplace. So we can't tell whether that was a valid kind of decision.

So the methods for validation turn out to be very important, and at this point are not in place.

And, finally, the future. I had a picture, too. This is a starry-eyed, wonderful future, right? Okay.

Computerized communication, no handwriting problems. I'm putting the next speaker out of business.

There will be new problems, and those need to be monitored, and we need to figure out what kinds of additional problems they introduced, and in the FDA's recent Web site the report for consumers,

there's a discussion of some experiments in hospitals with bar codes to permit computer reading of prescriptions, which if it expands will be presumably a way to reduce error.

Unfortunately, the future isn't here yet, and in the meantime we need to proceed with caution, and I wish you all a lot of good luck on a very difficult problem.

Thank you.

(Applause.)

DR. COHEN: Thanks a lot.

And we had two questions for Kaz Jaszczak.

How much handwriting distortion is appropriate to reflect the real world?

Three questions. How much verbal distortion is appropriate to reflect the real world.

And how about errors in E-prescribing being anticipated and evaluated?

So our next speaker, Kaz Jaszczak.

MR. JASZCZAK: Good afternoon. My name

is Kaz Jaszczak, and I'm Director of Product Planning and Operations at Parascript.

And I'm going to talk today mostly about evaluating drug names, similarities, applying handwriting recognition technologies.

Also, I would like to touch a little bit about speech, let's say, recognition systems, how they can be applied, but in general I will put more attention to handwriting because I think that handwriting technology is much more advanced, and as it is right now it can be at least partially used even right now to the tasks which we are talking about today.

In addition to that, the company I am representing is specializing in handwriting recognition. So we have pretty good experience in that area, and we think that the techniques which we develop for recognition of handwriting can be also used for determining similarities of handwritten words.

Okay. So the goal is to evaluate

preparatory drug names to reduce medication errors due to similarity in drug names, and the optimistic answer is that Parascript's technology can compare graphic patterns of writing, a proposed drug name against the patterns of written (phonetic) the existing drug names.

DR. COHEN: Excuse me, Kaz.

MR. JASZCZAK: Yes.

DR. COHEN: Excuse me just a minute. We're having a little bit of a problem as you can see with this screen kind of jumping. So what we're going to do is adjust it, and it's going to take about three minutes.

MR. JASZCZAK: Okay.

DR. COHEN: But we are okay for the time. So I hope you won't mind. Just hold on, please.

MR. JASZCZAK: Sure.

(Whereupon, the foregoing matter went off the record at 1:38 p.m. and went back on the record at 1:41 p.m.)

MR. JASZCZAK: Okay. I think we are

ready.

Okay. Maybe I will tell just a couple of words about Parascript, not that I would like to promote the company, but I would like to promote the technology, and Parascript is a recognized industry leader when it comes to handwriting technology. We have about 20 years of experience when it comes to handwriting recognition and some of you probably remember the Newton device in which this technology was first time deployed, and after that it was significantly improved.

So Parascript was the first company who introduced handwriting recognition, and right now we recognize more than 100 million forms a day, and by forms I mean real forms, mail pieces, checks, et cetera.

Our technology was developed a little bit different way than other people do. Usually people start with OCR, which is optical character recognition, which usually is limited to machine print.

We started with the most ambitious task, which is handwriting recognition, and our approach was to recognize things like human being is doing that. So very often when you are looking at writing, which is kind of free writing, you are not able actually to segment this writing into particular characters. this is what you usually can do on machine print.

In that case, simple application of neuron (phonetic) networks on character level can lead you to very good read rates on character level. Like human being is reading handwriting, usually you are reading things on word level, and we are applying a lot of additional knowledge to recognition, and this knowledge is different type of context, like dictionaries, type of templates, some syntax, semantic information, et cetera, et cetera.

So Parascript developed specific technology for describing any type of words with a combination of descriptive language, and this descriptive language simulates to some extent motions

which we are doing when we are writing things and strokes which we are making when we are writing.

I'm showing here on this slide this type of elements of this language. We call them XR elements. So these are minimus, maximum, some different shapes of curves, and we have 64 actually, this type of different shapes.

So when you are looking statistically at all types of handwritings, each word can be represented with this kind of set of strokes, independently if I'm writing this or any of you is writing this.

Of course, our writing style will be different, but you know, the basic combination of certain elements in your handwriting will be similar to my handwriting.

So here, for example, this letter D or CL on this particular picture, this is how it's represented with this little elements which you are using, and as you can see on the bottom, this is a very good example when I wanted to talk about context

and how it is important for recognition.

You are having two words, "clear" and "dear," and obviously if you would like to recognize this only on a character level, even if you can afford perfect segmentation in this case, you don't know which segmentation is correct because both segmentation with CL and segmentation with V at the beginning are correct segmentations.

So you can come with the same probability, let's say, or confidence level for your answer on both words. Only when knowing, you know, what is the context or what is the dictionary in this moment you can tell that this is one of these words.

So we are doing two types. We are actually having a number of different engines, but two basic engines. It's kind of analytical, taking this analytical approach when we are generating this kind of set of this funny signs which you see on the screen. And whenever we can segment into characters, we are also using neuron networks.

When it comes to similarities of names,

of course, for recognition what we do, we don't only look for similarity. Actually our output for what we recognize is usually a confidence level, but on the way of recognition, we definitely look for some similarities, similarities of what is written versus of what our knowledge is, the knowledge built into our engine, how things should be written based even on as key, let's say, representation of effects.

So we propose here two approaches which might help with at this initial screening of new names. The first approach is compare graphic patterns of writing of proposed drug name against the drug names existing in a database.

So the requirements will be as follows. We will need a set of patterns of writing, a proposed drug name, and minimum will be something like 50 samples received from different physicians, and we need obviously a database of existing drug names. So this is first approach, and we think that this is a more feasible approach.

The second approach, which will be

probably more efficient approach, but I don't think it's feasible, is to compare graphic patterns of writing a proposed drug name against the graphic patterns of drug names existing in database.

So in that case, we will need to have two things, the same thing as we needed in the first approach. So we will need to have pattern of writing a proposed drug name, but also we will need to have a database of graphic patterns for all existing drug names.

So that I don't think that this type of database exists, and to build this type of database, I think this would be pretty cumbersome task. Obviously, if you will decide even for the first approach going forward, we can start building this database at least for the incoming names. At some moment we can kind of switch to the second approach, but I think that, you know, it is more feasible to start with the first one. So I will maybe describe how the work flow will look like in this first type of solution.

So when the new drug name comes and let's say that this is Excedrin, okay? So we are gathering this, let's say, limited number of samples for this particular name. We are doing feature extraction, and this feature extraction is actually parameterization of all the samples which we gather.

So we are building a table of different representations of this particular name, and at the same time we are having a dictionary or database of all existing drug names as we have them now, and we are able to generate handwriting prototypes based on the power of our recognition engine, which correspond to similar, let's say, sets of combinations of this XR elements.

And we are applying fuzzy logic which kind of compares these two, and we are coming with a similarity score. So we can actually sort for you all existing names versus the samples which you acquired for the new coming name.

Now, what type of distortion we allow. Actually Parascript technology deals with any type

and quality of writing, including sloppy handwriting, and we do not limit writing to any writing style. So we accept any type of writing style.

In a while I will show you a couple of different forms on which we are doing recognition so you will realize that we are actually covering any writing style, and we do not require any training.

Obviously when the samples will be required, it is good to have diversity of different writing styles and maybe even taking some time in some separated period of times because people never write things the same way.

We also are having a product for signature verification, and when we were gathering data for forging actually signatures. So we had, let's say, a kind of reference signature, and we were trying to generate forged signature. We gave this to a number of people to do that, and each time it was different, but it was different each time they were doing this, but it was also different when they were doing after a couple of days.

The same with your signature, for instance. When you are signing your check, look at your signature after one year, and you will notice that there is quite different signatures.

So, you know, when I'm talking about collecting the data here, the data, it is good if the data is collected, let's say, in some separated intervals of time.

We also provide mechanisms for looking for similarities, also VR recognition on different representations, let's say, of the same word. So we call this an alias mechanism. So you not only can provide, let's say, dictionaries, but you can also provide aliases, and aliases can be simply replacement of the word with some nickname or it can be kind of an abbreviation of a given name.

And our technology also probably will be very helpful for looking for a similarity of mistakes, which are made in particular handwriting, which are the results of misspelling. So, for instance, if you are, let's say, missing some letters

or if you add some additional letter or if you will switch the order of particular letters in a name, it will also quantitatively measure, you know, how close this typical, let's say, misspelling is similar to existing names.

So here I just would like to support what I've said about, you know, this kind of independence. I would like to show you a couple of examples of forms which we are able to recognize, and this is one of the legacy forms on which we are able to recognize not only handwriting, but also some, let's say, symbols, and this is put independent to forms filled out with pencil, forms with condensed lines, things like correction, et cetera.

Regarding writing style, this supports actually the statement that we are not dependent on anybody's writing style. Parascript technology is used by USPS for mail sorting, and obviously we have to be prepared to read any type of writing, anybody, let's say, in the United States. So we sort almost 100 percent of USPS mail for bar code spraying on the

bottom of the envelope. When you are seeing this bottom bar code, this actually sprayed based on Parascript technology which recognizes the address and generates all of the information about the address.

And the last form which I would like to show is the form from I think it is 1910 census, and with it the pilot here. This applies a little bit. I think it illustrates a little bit, you know, also your needs because this is used for search purposes, and this is for LDS charge, which has huge genealogical archives, and they are going to provide a kind of automatic search service.

Because, you know, looking for some names in all of these archives takes weeks to months sometimes, and with Parascript technology, you can kind of index particular fields by image representation of these fields.

And now I would like to switch a little for you. This demo is based on our technology of searching actually, but I would like to kind of show

you this concept which I was talking about.

Let's say that we have sample. This is the second approach which I thought is less feasible, but we obviously can do the reverse thing as well.

If we have a list of different names of medicine and we have the database of snippets of names of this medicine, so now with this technology what I can do, I can generate a query for particular names. Let's say that this is this name. It goes simply through this little database, and I'm hearing just 50 entries here for demo purposes, but this can be, you know, a pretty huge database, and it looks for the best match to that query which I generated in the text form.

So I actually type this in, and it analyzes all of my snippets which I have in the database, and it lists them in the order of similarity.

So as you see, the two first entries are corresponding to what I typed in. On the third position you are having the name, which is very close

to the first name.

For instance, if I would do, let's say, some misspelling in this query, you can see that the first one still comes, is first, but as the second one I'm already having this different name. So I can very quickly and easily review all of this database and give you some candidates which are suspicious. Okay?

And so I can do kind of initial screening for you, and later you can also apply all other criteria for the name which you guys talked here about, but you already have kind of filter at least with handwriting.

Of course, like I'm saying, in this case we will need to have a pretty big database of written names which already exist, but the technology allows also to do the opposite. So we are able to do the opposite. We are able actually to have just samples of the new name and dictionary of all existing names. So this also can be done.

So I think this concludes what I wanted

to say about handwriting. There is much more to it, of course, but I think that I passed on the concept here, and I am very curious, you know, what will be the feedback on what I've said.

Thank you very much for this opportunity speaking here.

(Applause.)

DR. COHEN: And so now we have time to ask questions of any of the panelists. So please feel free to step to a mic, and let us know your name just before you ask a question so that we can record that.

Any questions?

MR. COHEN: Yeah, hi. I'm Bob Cohen from Lexicon. Is this on?

DR. COHEN: Yes, it is.

MR. COHEN: Lexicon Branding.

I want to pick up on a comment that Ms. Diamond brought up about the fact that products are not yet on the market.

When something is brought to market, it's

accompanied by detailing. It's accompanied by promotion. It's accompanied specifically by a lot of literature in medical literature so that it comes to mind for doctors and pharmacists both, as well as possibly for consumers.

And the kind of methodologies we're talking about here, that's not the case. People want to be right when they answer a question. So they see a name written out that is not familiar to them, even though they're told that these names may not be familiar.

Are we not building into that system innately high error potential?

And if so, how do we account for it?

MS. DIAMOND: Sure. Two things. One is I think you're absolutely right it may very well be that once something hits the market and it's surrounded by all kinds of other cues, that some of the things that you would detect in a premarketing stage would disappear. No question about that.

In terms of the kind of testing you can

do to control for people's sense that they want to say something is familiar, that's easily handled by having duds, you know, controls in the testing that we know are nonexistent in addition to the thing being tested and to test for that person's rate of just agreeing or that group's rate of just agreeing that they recognize something.

So that is handled, but that doesn't take away the issue that premarket is different from after something is in the market.

DR. STROM: If I can follow up on that, the other thing to keep in mind is unlike the normal research setting where your focus is on the mean and the average and will most people respond to the drug and so on, here you're looking for the outliers, and, yes, because you're looking for the people who are going to be making the mistakes, and, yes, there's a marketing effort, and, yes, they're detailing, and, yes, there's advertising.

That doesn't mean it's going to hit everybody, and the people who are most likely to be

making the errors, perhaps for the reasons you're implying, which would make sense, may be the people who didn't get to.

And so it may be that the premarketing and post marketing setting are certainly different in many ways, but they may not be different in this key way.

Again, that's a testable hypothesis, and the central point that I want to bring back is there's an enormous amount here which is testable and is researchable, and we shouldn't be going just on the matter of a question of faith and clinical subjectivity. We should be doing research in order to find out the right way to do these things.

DR. COHEN: Yes, ma'am.

DR. DORR: Bonnie Dorr, Department of Computer Science at the University of Maryland.

And I just wanted to address a point that Brian Strom brought up. I really like the talk, by the way.

I absolutely agree that if you have a

list of names that reviewers have already seen and then they're asked questions, they already have that data in their heads. So that's not really a valid gold standard.

But if you had such a list and, say, a bunch of systems were developed for drug name matching not based on that list and then you ran a bunch of comparative experiments on that as a gold standard, then I think that is a valid gold standard because the developers of those systems presumably haven't seen the list and haven't done judgments of those types. So that's just a comment really.

DR. STROM: Yeah, I certainly don't disagree with what you're saying at all. I think the more general point is we should be simulating reality in whatever way is practical and as close as we can in measuring actual observed error rates, not that that would replace what's being done today ideally, but that would be a gold standard by which we could evaluate.

I mean, the point is when there are a

dozens of different variations on the way things are being done the way we heard this morning, somebody is doing it wrong. Hopefully somebody is doing it right also. Hopefully multiple people are doing it right, and we need to be able to evaluate which is right and which is not.

DR. DORR: Well, I think you hit the nail on the head when you said we don't know what the question is yet. We don't know what we're testing. We don't know what our thresholds are so that even if we were to evaluate what we decided was a valid gold standard, we'd still need to know what the numbers need to be in order for it to be a possibly confusable name pair.

DR. COHEN: Any other questions?

Brian, it's certainly possible that what is being done now might even be the gold standard eventually.

DR. STROM: It's certainly possible that at least some of the things that are being done now are correct. It's also possible none of them, but

it's certainly possible that some where. It's not possible that all of what's being done today is correct because it's too variable, and the issue is what parts of what's being done is right and what parts are not right. That's what needs to be researched.

DR. COHEN: Thank you.

And, Kaz, one of the things that we run into with looking at names when it's done is it's not just name versus name, name confusion, that is, but also occasionally there is confusion with some hospital terminology or laboratory tests or, you know, other elements of a prescription.

And I assume that you could build that into your technology, which I think is fascinating.

MR. JASZCZAK: Yes. I think any additional information is very useful, and during plans I had actually conversation while I was giving example how you can strengthen recognition rates by using additional information, and in particular, in made processing, we are using not only recognition

itself, but we are using also behind this USPS database for cross validation of what we recognize.

So we are able to cross-validate different elements of the address. This way we come with much better answer, with much better read rates.

Similarly here, if you are looking at names, and if you can have brand for this, if you can add, let's say, dosage into this and if you can look not only for similarities between name itself, but also these other elements and if you can add cross-validation with different elements, definitely your analysis will be superior.

DR. COHEN: Any additional questions? I have another question here.

MR. HARTMAN: Steve Hartman from Novartis.

I had a question for Shari Diamond about focus groups.

I was curious whether there is any data on whether you can reduce the problem of interdependence by increasing to a reasonable size

the number of participants in a focus group so instead of having five you have 20 or 25.

And then the other question related to that was: do you need a moderator? For example, if you had in this particular case an expert panel of pharmacists and risk safety analysts sitting around talking about various different possible drugs, do you really need a moderator at all? So could you eliminate that element as well?

MS. DIAMOND: The typical focus group, at least what's been discovered works best for just running a group, is about eight to 12 people, that is, fewer -- and it varies depending on who's in the focus group, of course. Some groups are more disciplined than others.

When you get much larger, you surely need somebody to direct traffic. All right? And so those two are not unrelated just in terms of running a focus group.

The real issue on a focus group is the interdependence of the members. It's the advantage

of the focus group because people feed off one another, and they come up with new ideas based on what the other person said. But it's just that potential suggestibility that makes them not very good for testing how the general population would respond, whatever that general population is because the responses of some of the people in that group who voice agreement with somebody else in the group may have been produced, influenced, suggested by precisely that mention of the other group member.

So it's a good, quick read on some things, but it's not a testing device.

MR. HARTMAN: What I'm trying to get at is that it looks as if a statistical model would be very difficult to create that will be practical and affordable and is readily at hand, and so it looks as if we may be moving towards something like an expert panel playing a very important role in the name approval process.

And so what I'm trying to get from you is some suggestions as to how to structure the expert

panel in a way to minimize this sort of guessing and interdependence and moderator influence.

MS. DIAMOND: Okay. Before we go there, if that's the direction you think we're going, then we really do need to do some serious testing on the correspondence between these expert panel positions and the kind of testing that we would do with a larger population of less sophisticated folks because we don't know whether they are predictive of how the population at large would respond.

So I think that's the first step before you go there.

DR. COHEN: Any other?

(No response.)

DR. COHEN: Well, thank you very much, panelists. Thank you for staying on time, too, and we'll take a break now.

(Applause.)

DR. COHEN: And please come back and sit down by 2:30.

Thank you.

(Whereupon, the foregoing matter went off the record at 2:10 p.m. and went back on the record at 2:29 p.m.)

DR. GROSS: Well, that was very impressive. I'm glad to see you're so anxious to go on with the next session.

The last question really was a perfect segue to our first speaker.

I am Dr. Peter Gross. I'm the moderator for the session, and I'm also chair of the FDA's Drug Safety and Risk Management Advisory Committee.

The emphasis in this particular section will be on decision analysis tools, although there's a fair amount of overlap with the previous session.

The first speaker is Dr. Rick Shangraw, who is CEO of the Project Performance Corporation. He will discuss expert committees, which was the last question asked.

His company uses multi-disciplinary teams to help clients solve complex information technology and management and environmental issues.

I'm going to introduce the other speakers, and then when they each come up to talk, I'll point out the questions that they were asked to address.

The next two speakers will talk about computer assisted decision analysis. Dr. Bonnie Dorr is Associate Professor, Department of Linguistics at the University of Maryland. She is a specialist in computational linguistics, which uses computers to assess the similarities of words.

Dr. Bruce Lambert is an Associate Professor, College of Pharmacy, the University of Illinois at Chicago. He has published on how short-term memory to recall drug names is affected by similarity, familiarity, and frequency of exposure.

The next speaker is Dr. John Gosbee. He is Section Director for Patient Safety at the Veterans Health Administration, National Center for Patient Safety. He will discuss premarketing evaluation and decision analysis through failure mode and effects analysis, or FMEA.

Dr. Gosbee has published widely and has demonstrated to health care personnel the benefits of human factor engineering to redirect a care team's focus on redesigning the systems to prevent adverse events from recurring.

He has also published on performing proactive risk assessment in health care by using FMEA, as well as retrospective assessment using root cause analysis.

Through a multifaceted program at the Veterans Affairs' National Center for Patient Care, he was able to accomplish a 900-fold increase in close call reporting of high priority events.

Our final speaker is Dr. William Campbell, who is the Dean of the School of Pharmacy at the University of North Carolina in Chapel Hill and also a member of the Drug Safety and Risk Management Advisory Committee of the FDA.

He will conclude our session and discuss premarket risk management programs. Bill has published on the limitations of current methods of

risk communication, and he has proposed some potential solutions. He is interested in the evaluation of pharmacy data systems and epidemiologic investigation of drug use.

Now, to the first speaker. Rick Shangraw will address the following questions:

Is an expert committee necessary to review information from studies?

How many people should staff an expert committee, another issue that came up previously?

What credentials are important for expert committee members?

And should the expert committee meet in person, via videoconference, teleconference, or E-mail?

Rick.

DR. SHANGRAW: Good afternoon. I've got the chance to follow up on a couple of issues that were brought up today, which is good, and have a chance to extend into some new areas as they relate specifically to expert committees.

First, I'm going to spend just a couple of minutes framing the problem, and we've heard over the course of the day already today a number of different conversations about folks using expert committees or expert panels in this process of looking at potential drug name confusion, and so I'm first going to talk a little bit about where expert committees or expert panels could be used.

And then I'll be going to these questions that were asked in terms of how you might use them. And in that regard I'm going to be bringing in a lot of the research not just from the field of the health sciences field, but also from many other fields that really play into this question of the value of expert panels and expert committees.

And then finally I'm going to bring up two other concerns that play very well off of the last panel about really what needs to be done next in terms of thinking through the use of expert panels and expert committees. So I'd basically added two more questions to the list. So we really have six

questions we're going to talk about.

You've heard today a lot of discussion about using expert panels, and as you listen today in my mind you really found that people are using expert panels in really three ways to address this problem or to make decisions on this issue. The first way or manner that they're using it is what I sort of all the all in one process. In other words, we've heard some speakers talk about the fact that they're using exclusively expert panels as a way of looking for name or drug confusions, and that's sort of a cradle to grave type approach. So you convene the expert panel, provide them with a new drug name, have them go through a process, an expert panel process, and then come out with potential name confusions, and they make a decision on that single process.

Another use of expert panels which I found enlightening from many discussions this morning that we're seeing in industry as well as in FDA is they're using expert panels throughout the process and, more importantly, at the end of the process as a

way of almost assimilating or integrating all of the different studies that contribute to understanding a potential area of confusion for a potential name, and I call this sort of the clean-up position, right?

So the expert panels come in, take a look at all of the different studies that may have been done, prescription studies, verbal studies, handwritten studies, and then the expert panel sits down and assesses those and comes up with some kind of conclusion.

And then finally, you've also heard today the use of expert panels as doing one of those factor studies, one particular study in a suite of studies that are done as a way of trying to understand problems associated with drug names.

And so when we come to this problem about thinking of expert panels, we're really coming to it not only trying to understand how you might use it in this process, in the different places where you can use expert panels in the process, but then once you decide to use them, what's the best way to utilize

expert panels, and that's what we're really going to focus on in terms of answering the questions.

So the four questions which I'll go through which were just iterated were, you know, when is it necessary. Are they necessary to use them? How many should staff an expert panel? What are the credentials? And what's the media or format in which you should use expert panels?

I have the benefit of being a social scientist, and as a social scientist, we like to reach out and look at many different disciplines in the way that we try to solve problems. And this particular problem is one that has been researched by many. I'm not going to provide any original research today, although I've done some research myself in this area, which I'll be talking about.

Most of the research that you see in the literature, especially experimental literature, come out of psychology and sociology. There's a large and emerging set of literature on expert panels coming out of the legal field, a set of legal researchers

looking at, you know, quality of jury deliberations, looking at definitions of what is an expert witness and even plans to look at science courts, which are specialized expert panels to look at very highly scientific problems.

On the policy scientist side, there's a tremendous amount of literature about the value of expert panel in the area of forecasting, that versus quantitative methods, comparative ways of looking at that, and some of the research that I'll be presenting today comes from that area.

There's also work in game theory, particularly also in organizational behavior and theory as a way of understanding how groups act and interact, which plays off a lot of the psychology and sociology literature.

And then the other side that has been interesting to look in the literature is that there is also a lot of work being done in the health sciences area and a lot of work being done about the use of expert panels clearly as a way of looking at

appropriate necessary care. I'll be talking about some of the NIH work in this area.

And as I go throughout the presentation, I'll be talking about some different methods or approaches that people use to conduct expert panels. some of them are very generic and have been around for several decades, the Adelphi method, the nominal group technique.

Some have also been around for a couple of years, and they've been particularly focused in the health sciences area, work being done in NIH in the consensus development program, work that was originated out of the RAND Corporation in collaboration with UCLA Medical School to do some work in the area of appropriateness methods, which also comes out of the health sciences area.

So as you can already begin to see, even though I have very specific questions to answer today, it's a very broad topic with an awful lot of research coming in from a very large number of disciplines, and so my objective today is to

synthesize that.

Now, one thing you won't hear from me today is specific research about the value of expert panels addressing this particular topic, and that is are there potential confusions with proprietary names. There's actually not very much research in that area. There's some research that's related to that from an empirical standpoint. Dr. Lambert has done some work in that area and some others in the audience have here, too.

But there actually hasn't been a broad base of literature built up in this particular problem. First of all, it's a relatively new problem, and second, as you've heard from some of the speakers today, there just hasn't been enough focus on methods and approach and looking at them empirically enough to build a scientific base for deciding which methods are better or worse, and you heard some concerns in the last panel about a need to do that.

And so there's not a whole lot of

specific evidence here to relate to this problem, but a lot of peripheral evidence from a lot of different disciplines, I think, contribute to better understanding the problem.

So the first question really is if you go back to the charge here: is an expert committee necessary to review information from studies? So is it necessary?

Well, as you begin looking through the literature, the question isn't really one, first, of is it necessary. Actually the question that everybody tries to answer first is are they of any value, and in particular, are expert committees of any value in comparison to other methods you might use to arrive at a decision.

And here not surprising, anybody who sat on an expert committee can understand this, there's a lot of disagreement in the literature about really just the fundamental value of them. Part of that is because there are issues related to the ability to structure them in a consistent manner so that you can

understand the outcomes in a reproducible way, and we've heard some of that from speakers earlier in the day.

But the research has been growing in the area that if there's a consistent method used, that obviously it adds consistency to the decision making process. There is some empirical evidence that says if you use an expert panel it might be better than using just a single expert. In fact, there has been a whole lot of research about whether or not using single experts versus a panel of them yields better results.

And, again, not surprisingly, there's been mixed results there. But the other part of the literature that's very clear is that if you don't run the expert panel in a consistent manner and if you're not cognizant of the potential problems of an expert panel, it will absolutely produce systematic bias, and I'll talk in a minute about what kind of systematic bias you can expect from an expert panel and also talk a little bit about the ways that you

can control for some of that bias.

And then finally, just an interesting side note. A recent study just came up in the last year or so that if you let expert panels deal with things like numbers and letter and substituting them, which is kind of an interesting subpart of the problem we're dealing with here, and that is looking at orthographic comparison and phonetic comparisons, that actually if you gave that to a group as opposed to an individual, the group does a better job solving those problems, which is kind of an interesting side note there.

By the way, as you see as I'm going through here, I've actually identified the literature for those in the audience who are interested in the literature. I've got the references here. If you want a copy of that just stop on up. I'll be happy to give that to you.

So in practice though, when is it necessary to use an expert committee? And after years of research in the forecasting field, in the

social sciences area, in the psychology and sociology field, most people have come down to a couple of quick rules about when you might want to use an expert panel.

So when is a necessity of using an expert panel? The first rule is use an expert panel when you don't have good historical data. If you have great historical data and you can build a model off that data, then you might be better off building a statistical model than you would be using expert opinion.

Second, use an expert panel if events in the future are likely to invalidate or be very different from events, but if there's historical data, don't use an expert panel.

And finally, if there are issues of ethical and moral concern, use an expert panel.

So as we think about this as it relates to the problem of looking at drug name comparisons and potential confusions, we certainly have a case where we don't necessarily have a strong enough

historical base, at least one that is articulated in a way that we can use it for being predictive in terms of are there potential confusions for names.

Some of that is a result of having difficulty in reporting issues related to potential drug name confusions. Some of that is that there are data bases that exist but the proprietary is not open.

But in any case there doesn't seem to be enough of a historical basis for that. So that could call into the need for having an expert panel.

And it is also clear that in some cases that there are likely future events that are going to occur that would cause you to want to bring a set of experts in that could have at least some insights into those things that are happening, changes in the packaging, changes in dose administration, changes in branding techniques and approaches that may be coming into the future that we didn't have in the past and, therefore would call into the use of expert panels.

So as we begin approaching the problem,

the question of is it necessary to use expert panels, I would say it would be necessary to use expert panels, yes, in this case, and I'll get in a minute to when you might want to use them.

I'm not saying though that you use them exclusively, and I think that one of the things that we heard this morning that was really important was your heard most of the private sector organizations, as well as the FDA, talk about the fact they use multi-methods as a way of coming to, tackling this problem of potential drug name confusion.

And clearly as you look across the literature in other areas of the literature, particularly in solving complex social science problems using a multi-method or multi-factor approach is one that has been very successfully shown to at least yield better results over the long term than a single method approach.

So can you use expert opinion or expert panels? Yes.

Should you use it exclusively? My view

is no.

Okay. Optimal size of expert committees or expert panels. The literature here is fuzzy, but the literature basically shows that for a number of different reasons you might want to rely upon groups in the area of somewhere between five to ten people or 12 people.

Most of that research, interestingly enough, has come off of sociological studies where they just found that there were problems in using different size committees from an effective functioning standpoint, and there has also been a lot of study in the communication literature that if the group size gets too large, you can't have complex communications because it gets too hard to moderate the panel or the group in terms of having complex communication.

Large groups beyond ten or 12 have been shown to be useful in expert panels if you change the way they vote on the problem at the end of the problem.

So in other words, what I mean by that is that if you have much larger expert panels, and there have been some Delphi panels that have been run over the years that have been approximating 100 experts in a group, that they've shown them to be better than smaller panels if you allow the way you change the voting patterns and the way that they respond.

In other words, you don't want to have a unanimous vote when you get that large, but if you bring down the super majority and majority voting patterns, you can have some value at the larger panels.

But in practice here, and if we look again to some of the longstanding, active groups that use expert panels, the RAND-UCLA research, the work that's done at NIH, even work that's being done by a set of research in the nominal group technique, you can see a lot of similarity, and actually you heard it just a minute ago in answer to a question on focus groups. About the size of an expert panel should be somewhere between eight to 12, and you hear that

number regardless of where you look. You either hear it in practice or you hear it in some research.

And so my recommendation: eight to 12. That's about the right number.

Credentials for committee participation. So this question really was surrounding the issue of what credentials do you need to have to participate on an expert panel, and the research, again, across all of the different disciplines is relatively interesting.

First of all, and I found this to be reasonably interesting, the set of studies that were done a couple of years ago. The recommendation was that the experts need to have some baseline level of expertise, but you don't want them too expert, and that most certainly ties into a longstanding set of research that says if people in a group feel intimidated by other members of the group, they won't contribute as well to the expert panel.

So the results here simply say you have to establish a baseline level of expertise, but you

don't want a bunch of gurus in the room because if you do, you're likely not to have a good interaction in terms of a solution for a decision.

The second set of research really revolved around who should participate from a different perspective, and you've heard today already a recommendation which I strongly support that to the extent you do two things. First, you try to match the participants and the expert panel based on the likely users of the results of the panel, and as you heard today, people are talking about populating their expert panels with physicians, with nurses, with pharmacists, with patients, you know, some subset of groups.

What the literature says is to the extent that you can make that group multi-disciplinary that still addresses the core of the problem, you'll get better results than if you had a single disciplinary response to the problem.

Clearly, as I said before, participant status affects the dynamics of the group. It's

something you have to be concerned about.

And finally, there must be some justification for expertise. So in this context it's important to establish some baseline level of expertise. Now, there's no clear solution here given that there's a particular problem you're trying to address. First is what the literature has looked at, but clearly you have to establish that baseline of expertise before you can be a credible member of the expert panel.

In practice, what you find is -- and, again, recommendations from people that have run panels. We've run numerous panels over the years -- is baseline qualifications are important. Conflict of interest is incredibly important, especially as you begin to look across the problem set that you're looking at in terms of potential conflicts of who can participate in the panels.

Domineering personalities is a concern, and finally, the concern about diversity.

The fourth question was what's the best

way to conduct these panels. Do you want to use E-mail? Do you want to use chat sessions? Do you want to have collaborative computer environments? Do you want to do it face to face?

And this is probably the area of research that has been most explosive over the last couple of years, given the acute interest of most researchers on whether it's better to be holding Web based or computer based group facilitated sessions versus traditional face-to-face sessions.

And here most of the research has tried to look at really three factors. Are decisions better when you use computers versus face to face, and on what set of computer mediated settings or groups are you more likely to get better results?

Are the folks that participate in these expert panels, are they more satisfied with the decision or what I call decision commitment depending upon what media or form they use for making a decision?

And finally there's a concern about media

richness, which is: to what degree are they able to address complex or simple problems depending upon what kind of computer media they use in trying to collaborate in these group kind of settings?

And, again, the literature, it's relatively new. It's beginning to focus up a little bit, but basically there's a subset of researchers that have found that computer mediated systems do, in fact, decrease overall effectiveness of group processes, especially expert panel processes, but that can be improved if the people that participate in the expert panel know each other.

Interesting, right? So, in other words, if you're using a bunch of folks on an expert panel that don't know each other, you're going to have poorer results than if you use a computer mediated type operation or type setting where the people in the group have some history with each other.

It's also clear that more complex communications occur over systems that are more rich in terms of the way they can interact. So computer

mediated or facilitated sessions where you have a lot of ability to interact versus simple chat systems or simple E-mail systems are less likely to be effective.

Adding audio to systems and video to systems particularly enhance their effectiveness, which has been a pretty consistent finding over the last couple of years, and in part that's because it improves the media richness. Remember it allows you now to have more complex discussions.

And then research I've done over the years on commitment shows that if you use computer mediated processes, participants are less likely to feel committed or satisfied by the processes and face to face.

And then finally, after an extensive meta review, meta study of the problem, after reviewing 200 studies recently, actually 80 percent of the studies that we reviewed basically said on balance we really can't find the difference between face to face versus computer mediated, collaborative technologies

along the line of decision quality, and so again, it sort of opens up the box with where is this research really going if, after the last decade, we can't find any real differences. Where are we heading?

So from a practical standpoint, it seems to me we're back into the same old mold that we've talked about earlier, and that is if you have the opportunity and the money and the feasibility, try to combine these techniques again. You're likely to get better results from combined techniques.

And you'll actually see in some of the current expert panel techniques that are out there that they'll start with a computer mediated discussion. In other words, they'll have a Web based board that allows you to get the initial question out to the participants and have them give some results back from a computer mediated forum, and then they'll move to a face-to-face forum to discuss the collaborative results from that initial computer mediated conference as a way of bringing in face to face interactions.

And some of the more advanced expert panel, expert committee methods and approaches are combining now both computer mediated techniques on the front end and face-to-face techniques on the back end to take advantage of both qualities of both of those approaches.

And then the second and probably more relevant point here is given the complexity of the problem that you're trying to solve here in terms of name confusion, it's clear to me that using lower computer mediated systems, E-mail, chat systems, you're not going to be able to get the richness and complexity of communications necessary to look at these problems. You're going to have to increase your way up the scale in using more of the advanced collaborative group technologies.

I'm sure some of you have sat in some of these new technologies out there now where you sit in the room and you vote by buttons and they have consensus building. You can see the results on the screen about how everybody is voting. You have

second, you have iterative rounds, and clearly, we're moving into an era that's going to be changed dramatically over the next decade in the ability of the technology to support those expert type decisions

The two last questions I'd like to cover are two that I thought I was going to get on this panel, and that is how do you address the classic problems that you see in expert panels associated with a concept called "groupthink." It is a concept that was developed by a researcher, Irving Janis, back in the '70s, and Janis basically said that in any group setting you're going to have a potential for especially an expert type panel type setting, a potential for folks to move towards the majority decision and then to have the group begin to collaborate and continue to think that that majority decision is the correct decision because people are less likely to voice any kind of dissonance or any kind of conflict based on that majority position, and it's basically called the theory of "groupthink."

And I've given you here some of the

current thinking on how you moderate panels in terms of making sure facilitators are impartial, making sure you assign the role of critical evaluator to all committee participants, making sure you rotate through a devil's advocate position for people on the panel, subdividing the panel to account for differences in facilitation and then give people second chances even on preliminary results of panels as an opportunity to look at options or non-majority thinking in panels.

And much of the computer mediated approaches that you're seeing being built today build in a lot of this thinking into their systems in terms of anonymity, in terms of voting on problems, and the way that you can respond and get feedback from systems.

And then finally -- and this is probably the most important point that came out of the discussions in the last panel -- and that is if we're ever going to be able to really ascertain the value of these expert panels, there needs to be first a

method that's consistent in the way that the panels are applied, and one of the things I haven't heard today is a lot of discussion about how people actually deploy their expert panels.

I hear they have them. I hear they populate them with experts, but we haven't heard a lot at least today about how they structure them. And structure becomes very important to the first issue that was addressed in the earlier panel, that we have to have a consistent method in the way you structure.

After method, you can then figure out if you can reproduce. After you reproduce, you can figure out if you can validate. But we have to start first with method.

And so for those that aren't familiar, there are a number of very well defined methods for expert panels, and I think the objective here first off is to make sure that there's some way of embracing some standard approach. I've just put up on the screen one of them. It's a nominal group

technique. There have actually been a lot of variance in this over the years, and as you move towards putting in systems you actually can reproduce and replicate and then validate, there needs to be some consistency in the front end in terms of process and approach.

So where is this all heading? On the expert panel side much of the research we're seeing coming out now in the foreseeable future is really going to be focused on the value of the computer moderator, computer facilitated side.

And then the last bullet, which I think is most exciting in this field, is there has been really a lot of work now being done on combining the use of expert panels with more empirical or data driven models as a way of trying to come to more consensus in particularly complex problems like you're facing today.

And as you look at, for example the FDA process which is trying to take a computer driven model which you're going to hear about in a minute

from the next couple of speakers and then combining that with expert panels, you're going to be able to learn more from the literature and other disciplines about how they're trying to take those two disparate approaches, qualitative and quantitative approach, and bring them together into a better decision making framework.

Thank you.

(Applause.)

DR. GROSS: Okay. Thank you very much. Thank you very much, Rick.

The next two speakers will talk about computer assisted decision analysis. The three questions that they should address are:

How can computer resources be used to objectively measure differences between name pairs, for example, at a distance, bigrams, trigrams, et cetera?

Number two, how can computer resources be used to calculate weights for various elements in name similarity, mitigating issues, and aggravating

issues?

The third question is can computer assisted pattern recognition support the decision process to determine name and name similarities?

Dr. Bonnie Dorr will speak first.

DR. DORR: First, just one correction. I'm from the Department of Computer Science, not the Department of Linguistics. However, I am a computational linguist, and I have an affiliation with the Department of Linguistics.

So this is actually kind of a joint presentation with my colleague Greg Kondrak at the University of Alberta, who is also in the Department of Computer Science, but he is also a computational linguist, and we look at problems of pairing up different strings for other purposes. So I, too, am an outsider, as most of the people in the afternoon have been saying they are, to the drug name matching arena.

So one of the things we do is we look at different languages, say, English and French, and we

try to match up words that are cognates of each other, and that's where some of this technology originated that I'll talk about today.

So the first question: how can we use traditional techniques for validating or not validating, but comparing drug names?

We'll look at a few approaches very quickly just so that you'd sort of know what's been out there for a number of years from the '60s and prior to that. String matching to rank similarity between strings, in this arena drug names, has been around for a while. There are two classes of string matching techniques. One is orthographic where we look mostly at spelling, and the other phonological where we care about sound to a certain degree.

And there are also two different methods of matching. So you have two different dimensions that we're looking along. One is orthographic versus phonological. The other is distance versus similarity where in distance we care how far apart are the two strings, and similarity, we look at how

close are the two strings. So those are sort of two sides of the same coin.

And just to look at those two dimensions up against each other. All right. So orthographic versus phonological, the first bullet under orthographic, you have distance, metrics. For example, the string edit distance. It also has another name, the Levenstein distance which you've heard about before, would compare what pieces of the string differ. So with Contac and Zantac the pieces that differ are the C-o and the Z-a. That is, about two-sixths of the string seems to be different.

So you look at how far apart they are, whereas with similarity metrics, like the longest common subsequence ratio, or DICE, which you may hear about also, they look at the string as the same.

So for Contac and Zantac the piece that's the same is n-t-a-c, our about four-sixths of each of the strings.

You can also look at bigrams, that is, two character sequences, or trigrams, three character

sequences, and that's where you get measures like DICE, where we're looking at how many of those bigrams are the same in the two strings.

And so I gave, again, the examples of Contac and Zantac under similarity. About half of the bigrams seem to overlap, and so that's the DICE metric.

Under phonological, there are number of phonological approaches. Under distance I've listed Soundex. It also has its cousin, Phonics, which Soundex looks primarily at consonants, in particular the first four consonant sounds, and tries to see how similar they are using something like the strong edit distance, actually a combination of two different things, phonology and orthographics, and then assigns a score.

Phonics actually does a mapping prior to doing that matching that's similar in nature as well.

Under the heading of similarity, phonological similarity metrics, I have the ALINE approach, which is what I'll focus on primarily

today. So the traditional methods that we've seen for decade are listed there under distance similarity, and orthographic and the distance under phonological, but sort of a newer alignment style approach which doesn't talk about distance per se or similarity in the traditional sense is the ALINE technique which I'll talk about.

It actually looks at every character in the two strings and has a weighted alignment technique for deciding how similar they are.

Okay, and these are just some more examples to flesh out what I mean by distance versus similarity, where the bottom line is when you talk about distance, if two strings are similar, you want the distance to be small. If two strings are similar, you want the similarity to be big. So for the "hordes" and "lords" example that I have up there, you could imagine counting the number of operations to convert one into the other. That gets you the distance metrics, replacing H with L and deleting the E gives you a number two.

With similarity, you count bigrams, for example. So OR and RD are in common with "hordes" and "lords." So you get a similarity of two.

In Example 2 up there with "water" versus "wine," the distance, of course, would be further. There are more replacements and deletions. The similarity in that case is zero. There are no bigrams in common.

So that gives you a feel for the type of thing we're looking at with distance versus similarity.

You can also compare the two. There's a formula that relates them and gives you some degree of analysis of the two different types of scores against each other. And again, for distance, string edit, you count up the number of steps it takes to transform one string into another. That's the example I already showed.

Often we divide by the length of the longest string to get a ratio instead of just getting a number like two or three. All right. So two-

sixths and three-fifths for the two examples above.

Okay. Now, to similarity. What we do for the longest common subsequence ratio is divide the length of the longest common subsequence by the length of the longest string.

So if you had "reagir," if that were a word, and "repair," the longest common subsequence is "reair" and the similarity score would be five over the maximum of the length of the strings, which would come out to .83. Whereas with something like DICE, another similarity metric, you double the number of shared character bigrams and divide by the total number of bigrams.

So for the same example, you would get the ratio that's shown at the very bottom. Again, the details are not important. Essentially what you're doing is you're comparing pairs of characters that are adjacent to each other, and you get a score in that scare of .4.

Okay. So I went quickly over the orthographic approaches. I want to move quickly into

the phonological matching approaches. I mentioned a distance based phonological matching approach called Soundex, which has been around for a very long time, and I also want to move into a newer technology, similarity based phonological matching called ALINE.

So for Soundex and also its cousin Phonics, what you have is a table of codes. You group letters, in particular, consonants, together into classes and assign each class a number, and then you map each word that you're trying to compare into some sort of number sequence where actually the first letter you keep the same, and then you add in the letters.

So "king" and "khyngge" with those two spellings reduce to the same string K52, and in fact, we're only allowed to -- in the traditional Soundex technique you're only allowed to take the first four consonant sounds.

So you get "knight" and "night," the two spellings reduced to very different strings, K523 and N23. Whereas "pulpit" and "phlebotomy," which are

very, very different strings, reduce to the same string of P413.

So obviously there are issues with the traditional phonological distance approaches. What went wrong? Well, for example, we truncated the word to four characters. We ignored vowels and used numbers instead of decomposable features, and I'm going to get into what I mean by decomposable features next.

Okay. Another possible approach. Say you were told to do some sort of phonological mapping. One approach might be to compare syllable counts or initial and final sounds and stress locations which would allow you to identify certain pairs, like "aloxi" and "floxin," but perhaps miss pairs whose stress patterns are different or a number of syllables are different, like "strattera" and "avantera" and "instrinsa" and "intralipid." So you might be missing pairs that you might otherwise get if you used phonological features to compare two words by their sounds.

Okay. What I mean by that, an example shown here. If you had X as a final letter of the word -- so that's what the X pound sign means. The pound sign is simply a word boundary -- that reduces to a set of features that I'm not necessarily going to go into, but consonantal is one of them. It's simply a consonant, alveolar, stop, and minus voice. This just gives information about how you are articulating the sound, the place, the position of your tongue, the type, the manner in which you're articulating the sound, whereas X at the beginning of a word sounds like a Z. So it has a different set of features or different positions of your tongue, and so on.

If you could break down the characters into these phonological features, you would perhaps weight the features according to what's important for the particular application that you're working on to get a better matching.

So phonological similarity reduces to an optimal match, finding the optimal match between

phonological features. For example, with Zantac and Xanax, you want to line up the characters according to how they sound.

And this is where my colleague Greg Kondrak comes in. He builds the ALINE system. Two fundamental components of ALINE are that it has a similarity function that uses linguistic features based on salience. So there are features like alveolar and stop. Alveolar just tells you where in your mouth you're articulating. It's actually just behind the teeth, line T in "tuh" and "duh." All right?

And stop tells you that there's a cessation of sound immediately following the character sound, and those are more salient than, for example, the plus voice feature.

And then the other fundamental component is that there's a method of choosing an optimal alignment. He creates the alignment based on a weighted multi-feature analysis, which I'll show in a moment.

So it's designed originally to align phonetic sequences for cognate matching. I mentioned that we're doing things like comparing English words to French words. So there's a lot of different computational linguistics applications that this would be useful for.

And I gave an example there, "colour" and "couleur," all right, for two different languages, but you would want to apply feature weights that are fine tuned for your specific application because the weights that you have for that task don't necessarily apply to drug name matching.

The approach is also efficient. It uses a dynamic programming algorithm to search for the correct alignment of characters in the strings. These details here aren't important. The top two, place of articulation and manner of articulation, are the highest weighted features. These are not binary features. These are multi-valued features. So within each one, for example, in the place of articulation, you could have a bunch of different

values.

So for bilabial, the first one, the very first picture, I have to show you the head with the teeth and the tongue and so on because that's the linguistics part of computational linguistics. So bilabial is where you've got the two lips together. Alveolar is where -- I think it's the fifth picture down -- where you've got the tongue just behind the teeth, and so on, and each one of these may have a different weight depending on what you're trying to do. These were the weights that were set for cognate matching.

The manner of articulation I'll breeze by also. That's just the way you're doing it. Are you stopping as you say the sound? All right. That would be something like a "puh" and a fricative. That is, is there some sort of vibration, as in "thuh" or -- sorry -- in "fuh" and "vuh" but not in "thuh." All right, and those also have numerical values.

All right. Addressing the question of

how would we weight certain features of the system, I'm actually talking probably at a much lower level than other people are when I say that I'm weighting features of the system.

There are different weights we want to apply to this problem for drug name matching than we would for cognate matching. We want to calculate weights for drug name matching based on a hill climbing search against a gold standard, all right, and we did tune parameters for the drug name task. Actually this is kind of a late breaking result. I didn't know about the USP list until last week. So we decided to run that through the system, and I'll show you what we got.

We also adjust other parameters like the maximum score that is cognate matching, allows you to have letters that are very far apart match, like "puh" and "kuh," but whereas that's not appropriate for the drug name matching task.

We also have a heavier insertion/deletion penalty in the drug name matching than we did in the

cognate matching. We also penalize if vowels are not sounding quite the same. We're penalizing much more than we do in the cognate matching task and also some of the phonological feature values are tuned.

These are just some examples to show you that running it on just a small sample of drug names we do get that Zantac and Xanax score higher than Zantac and Contac, for example. Whereas with edit distance, LCSR, and DICE, we don't get that ranking. We get Zantac and Contac ranking higher than Zantac and Xanax.

All right. So that just gives you an idea of the types of distinctions we're getting when we run these. Our evaluation, as I said, is against the USP quality review, March 2001. In fact, we in there found 582 unique drug names. There were 399 true confusion pairs, according to what's listed there, and again, we don't know where those confusions are from. They might be from something other than whether the drug names sound the same or look the same.

But because this is what we had to work with, that's the gold standard we used.

There were 169,000 possible pairs if you sort of do a combinatoric permute on all of these. All right. So you could get a lot of different pairs out of this list, many of which, in fact, almost all of which are not part of the true confusion list.

All right. So what we did was we ran the systems through, and this is just showing you what DICE gets where "atgam" and "ratgam" actually got the highest value, and there's a plus next to it, meaning, yes, that pair did occur in the confusion list. These are just the top several.

All right. So herceptin and perceptin also scored very high, whereas the next one down is a false positive. It did not appear in one of the 399 pairs on that list that was publicly available. In fact, we think this might be a typo, and if you go look at the list, those two are actually in there.

The next one down, quinidine and quinine is also a pair, and so on. All right, and again, the

next one is a false positive, didn't appear on the list. However perhaps this method, the DICE method which actually we're not using, but it's a fairly decent method, is finding something that could be confused there with just the dash, U at the end.

This is a graph showing everything from ALINE down to DICE. DICE is actually the lowest scoring one. The edit distance is the next one up. Again, that's like the Levenstein distance. LCSR is the green one in the middle.

The pink and the blue top ones, those correspond to ALINE. The top one is ALINE without phonetic transcription. So even if we run ALINE without transcribing the string into other characters that are phonologically relevant, we get the pink line, which gets an average precision of .36.

By the way, what does this graph mean? The Y axis is the interpolated precision. That tells you out of all the ones we got at a certain recall value, that is, the top, say, 100, all right, how many of them were correct. All right, and the recall

tells you how many out of the total number of 399 pairs did we retrieve.

So although we don't really have a threshold here, this graph tells you you could threshold it at any of these values. If you don't care that your precision is very low, you could decide to recall all of them at the end of the chart, or you could do -- if you want a higher precision, you might only want to recall, say, 20 percent.

All right. So this gives you what is called 11 point interpolated precision, and then you can take an average across all of them, and I've got those averages listed up in the box there where ALINE is .36 at the top and DICE is .27 at the bottom.

Just to make sure we weren't fooling ourselves by having phonetic transcription do most of the work of the ALINE technique, we did apply phonetic transcription prior to running DICE and LCSR, and we ran the experiments again, and we got that ALINE is still at the top and the other two, LCSR and DICE, really minimally changed with the

phonetic transcription in their values. The averages came out to be the same across the interpolated precision.

So concluding remarks. So experimentation with different algorithms and their combinations against a gold standard might lead us toward some standardization of techniques that we want to do for evaluation. ALINE has a strong foundation for automating minimization of medication errors, we hope. This is something we would like to investigate.

We do allow for fine tuning based on comparisons with the gold standard. We can reweight the phonological features, and I mentioned that a little bit earlier.

This is related to pattern recognition. So the third question was about using patterns recognition techniques. In fact, when we run the ALINE algorithm we can discover patterns of predictable matches based on feature values. So we may discover that bilabial is in a very important

piece of a pattern to match when you have two drug names. So that part of the pattern has to match, whereas plus or minus stop might not be as important. So you can discover these patterns.

And that's it.

DR. GROSS: Thank you very much.

(Applause.)

DR. GROSS: Dr. Bruce Lambert is next and will address the same questions.

(Pause in proceedings.)

DR. GROSS: Everyone is right on time. So we're still in good shape.

DR. LAMBERT: Sorry about the brief delay.

I want to talk about the same set of issues that Bonnie Dorr just finished talking about, and, in fact, Bonnie talked about several of the things that my research has been based upon for the last several years. So I'll probably skip over some of that to avoid redundancy.

So the overview. These are the questions

we were asked to address. You are already familiar with those so I won't dwell on them.

I want a brief preface. We need to change some of the focus of what we're talking about, and we've addressed this already to a certain extent. Drug names, it's not enough to focus on drug name. We have to focus on drug products.

We have to keep our laptop plugged in, too.

What I mean by the difference between names and products, a name is just a name. When I refer to a product, I'm talking about all of the other attributes of the product, the strength, the dosage form, the route of administration, the color, the packaging, the storage circumstances, et cetera.

A similarity is not enough. In fact, similarity may be the least important thing. Frequency is a much more powerful driver of errors than similarity.

In fact, there's a guy names James Reason who wrote a very famous book called Human Error, and

he called similarity and frequency the two fundamental mechanisms of human error, and that's because we're so biased that frequency biases our perceptual judgments and our memories so strongly that you have to consider frequency when you're thinking about patterns of error.

And in the context of drug name confusion, it's prescribing frequency.

Also, error reduction is not enough. We need to focus on harm reduction. The vast majority of errors cause no harm. So we could reduce the error rate a lot, but if we don't focus on particular kinds of drugs, especially these narrow therapeutic index drugs, we're not going to reduce harm as much as we ought to. So I think the focus ought to be on harm reduction, not necessarily reduction of the pure number of errors, although that's obviously desirable.

And how do we balance this public risk against private benefit? We've addressed that already today.

So these objective measures of similarity, some of which Bonnie just discussed; these bigram and trigram, at a distance measures, I won't describe them further. They're all described in detail in a series of publications that I've written since about 1997, references to which you could easily find through Medline or by contacting me.

The N-gram and edit distance measures can be used on any formal representation of the name, either the spelling or the phonological representation of the name. Bonnie went over a lot of this again. So you can use the spelling of the name and look at bigrams and trigrams or edit distance, or you can use a phonological alphabet or phonetic alphabet like the International Phonetic Alphabet, which you'll see in a dictionary next to the name, or you can use something like the ARPAbet, which is what speech recognition researchers often use, and here I take the name Zyprexa and give you its representation in this particular phonetic

alphabet.

There are many, many variations of these basic measures, as Bonnie alluded to. You could add spaces before or after the names to emphasize the beginnings or endings. You could use different weights depending on the position of the letters. You could weight the different phonological features differently, a Bonnie illustrated. You can use different equations to compute the numerical similarity.

You could allow approximate matches between letters. For example, M is much more similar to N than it is to Q, and you could capture that fact. All vowels are more similar to one another than they are to any consonant, and you can capture that fact as well.

What's nice about objective measures? Well, they have lots of desirable qualities. One is they're a prefect reliability. You can compute the DICE coefficient this morning. You can compute it this afternoon. You can compute it tomorrow morning.

It will always be exactly the same, unlike expert committees and o on, which are notoriously unreliable.

Also they're very powerful even just for simple descriptions. For example, you can do things like compute the most common three-letter prefixes in U.S. brand names, which happen to be pro-, bio-, car-, tri-, vit-, pre-, nut-, ult-, con-, and per-.

I know some of my colleagues in the drug industry tell me that they just won't accept any name with any of these prefixes. So you can have simple descriptions.

You can also have simple descriptions of how long drug names are, how similar they are to one another on average. You can look at the distribution of their similarities, all of which I've done in this paper that I cite from the Drug Information Journal, and I think that adds. That gives us some reference when we're talking about, well, is this pair of names more similar than the average, less similar than the average. Where does its similarity score in a

percentile basis?

We can do all of those things using these objective measures.

The objective measures do predict the probability of human error, and I think that's the most important characteristic of these measures. And I've done a series of studies on short-term memory, visual perception, and comparing objective measures to subjective measures, which I think do validate this.

Most of what Brian Strom was calling for and a certain amount of what Shari Diamond was calling for I've already done, and you could take a look at this literature to see for yourself and evaluate whether or not these methods are validated.

When I say my methods are validated, what I mean is I've done the validation studies which show both the faults, the strengths, and the weaknesses of the methods, but I think that's a step beyond what's been done for most of these methods.

So similarity accurately distinguishes

between known error pairs and non-error pairs, for example. Greater objective similarity scores are correlated with higher rates of recognition memory errors by both lay people and pharmacists, and this is the example of the sort of line-up task that Shari Diamond described.

Greater similarity scores are correlated with lower rates of free recall errors. Now, no one has mentioned this today, but I published this in the American Journal of Heath System Pharmacy earlier this year, late last year. It showed that, in fact, the most similar names are actually easier to recall.

That is, if you know the name ends in "-statin" you can use that fact, Simvastatin, this-vastatin, that-vastatin, and you can run through your mental lexicon of all the statins and remember a particular drug which you may be trying to remember, and actually some people at USP have told me, well, that's what they like about generic stems, that they increase recall for generic names, and so on.

So similarity is not universally bad. It

depends on the task. In fact, it enhances recall very, very clearly.

Objective similarity scores are correlated with subjective similarity scores for both experts and lay people, and I've got evidence about that, and similarity scores -- come on back. Sorry.

Okay. We're going into the actual data now. So you'll have to check the papers to see all of the details, but this shows the relationship between spelling similarity and pharmacist errors. The citation is at the bottom of the slide. The slides will be available from the FDA at the end of the meeting.

But what it obviously shows is that up until a certain level, similarity has very little effect on recognition memory errors, but beyond a certain level, there's a linear relationship between increasing similarity and increasing recognition memory errors. That is, the more similar they get beyond a certain point the more likely you are to misrecognize a name.

This is the effect of spelling similarity on pharmacist free recall, and so we see the opposite trend. The more similar that the names get the easier they are to recall, and here we had pharmacists and lay people recalling simple three name lists of brand and generic names.

Now, this is the effect of sound alike similarity or phonological similarity on recall. It's not quite as straightforward. Phonological similarity actually does increase errors up to a point, and then as they really get similar you see the same effect as you do in spelling with greater similarity leading to fewer errors.

What's happening there is there's a rhyming heuristic. If you know that the name you're trying to remember rhymes with another name, you can use the rhyming heuristic to generate those names in recall.

Again, the details are available in this publication from Psychology and Marketing, but what this is is a graph that on the horizontal axis it

shows this trigram measure of similarity. On the vertical axis it shows a subjective measure of dissimilarity based on grouping names, similar and dissimilar groups.

And what you find is that although the relationship is far from perfect, the more similar the names are objectively, the lesser the subjective dissimilarity is, which is exactly what we would predict. And what this illustrates is that the objective measures are, in fact, strongly correlated with the subjective measures, which is what we want.

The next idea is a notion that no one has talked about before but is actually central to psycholinguistic theories of visual perception and auditory perception, and that's the concept of a neighborhood, and here we're not talking about, you know, where does Bob Lee live and what sort of neighborhood does he live in, but we're talking about the neighborhood of the drug name.

And there's a similarity neighborhood, and so you can think of the name that we're

evaluating as the target name, and that name will be similar to a certain number of other names. So within a certain distance we'll call that the neighborhood, and the number of other names inside that distance we'll call the density of the neighborhood.

What's also very important is the frequency of both the target name and the frequency of the neighbor names. These things are fundamentally important to how easy or difficult it is to accurately perceive a name either visually or auditorially. So there are these characteristics of the neighborhood, the frequency of the neighborhood, the density of the neighborhood, and the neighborhood radius.

And here I give just a simple graphical illustration, and this is from a paper forthcoming from the Journal of Social Science and Medicine about pharmacists' visual perception of drug names, which will be coming out some time towards the end of the year.

So here in the center you have the target name. That big star is a representation of a higher frequency neighbor. The little one is a low frequency neighbor, and the radius shows how big the neighborhood would be.

So here's examples of dense neighborhoods, high and low frequency. So you can have, you know, a high density but low frequency neighborhood where the target name is very, very commonly prescribed and the neighbor names are very rarely prescribed. You would expect that name to be relatively easy to identify even though it had a lot of neighbors.

In contrast, the figure on the right is a low frequency name with lots of high frequency neighbors. You would expect that name to be very difficult to correctly identify.

And here is all possible combinations of neighborhood frequency, stimulus frequency, neighborhood density, and I use these in this visual perception experiment to identify the importance of

these various factors.

So what are examples of high/low stimulus frequency, neighborhood frequency names? High stimulus frequency names are just commonly prescribed drugs. In the database I used, which was a government database, Ventolin, Dyazide, Provera, these were names whose log prescribing frequency was greater than seven.

The uncommonly prescribed names were things like Vistazine, Antispas, Protophane.

Names from a sparse neighborhood, a name like Flexeril, which in the National Ambulatory Medical Care database I could find no neighbors within an edit distance of three for Flexeril.

In contrast, you take a name like Dynabac and you find Synalar, Rynatan, Dynapen, DynaCirc, Cynacin, Cinobac. It's in a much denser neighborhood. So clearly we could already see that it's desirable to lace new drug names in sparse neighborhoods and to avoid increasing the density of existing neighborhoods.

So this is just one of many, many results from the visual perception study. Obviously the overwhelming trend here, on the horizontal axis is the frequency of the target name, the name that we're trying to identify, and what do we see?

As the frequency of the target name increases, the error rate, which is on the vertical axis, increases dramatically. This is the most fundamental finding in all of psycholinguistics. It's called the word frequency effect. More common words are easier to identify, and this task was a very difficult task.

We took typewritten and handwritten drug names. We superimposed a whole bunch of noise on them. We deleted a bunch of the background, and we only gave the pharmacists three seconds to identify the names. So this is a very difficult task.

And you can see that even in this very difficult task for the very common drug names, they were relatively easy to identify.

The difference between the blue and the

red line is the difference between high density and low density neighborhoods, and what you see is that in the blue line these are drug names from high density neighborhoods, and you find that in high density neighborhoods, just as we would expect, drug names were harder to identify, but the density of the neighborhood only mattered for low frequency drug names.

So for very commonly prescribed drug names, density doesn't have much effect, but frequency always has this very, very powerful effect.

So what do I conclude about objective measures? They work. They are not perfect. They are much better on a population basis than they are on an individual basis.

What do I mean by that? The analogy to smoking is the best way to explain this. I think most of us in this room, unless there are some tobacco executives hiding in the back, would agree that smoking causes lung cancer, but does anybody know what proportion of smokers actually get lung

cancer? It's less than 25 percent.

So 75 percent of the time when you use smoking as a predictor for lung cancer, you're going to be wrong. Seventy-five percent of those smokers identified as potential lung cancer patients are false positives.

Does that mean smoking is not a risk factor for lung cancer? Of course not. It just means that these things are difficult to predict, and even very good predictors, things which we would recognize as excellent predictors, like the relationship between smoking and lung cancer, are wrong much more often than they're right.

So these things, because of the nature of the false positives, they're much better for public health. So we could tell everyone to quit smoking, and on a population basis as they quit the lung cancer rate will go down even though lots of people who quit never would have gotten lung cancer.

So on a population basis if we decrease similarity we will decrease the number of name

confusion errors even though some of the names which we prevent from getting on the marketplace never would have been confused.

And in conclusion, we should be using these objective measures.

So I want to digress briefly into a demonstration of the software we've developed, which is described in a forthcoming article in the Journal of Medical Systems, which will come out at the end of the year, and in much more detail in our patent which was granted March 4th, this apparatus, method, and product for multi-attribute drug comparison.

So briefly I'll just switch gears and get out of my real one. Okay. So I've already run some of these searches. I ran them on Zyprexa, not out of a desire to embarrass one of our hosts, but because we already know that the Zyprexa has been confused with some other names.

So here we have just a trigram search on Zyprexa, which not surprisingly ranks Zyprexa as the most common name or the most similar name in the

database.

I realize for those of you in the back it will be difficult to see this. The second closest name using a simple trigram similarity measure on spelling is Zyprexa Zydis, which is an alternate formulation, I guess; Zyflo, Zyvox, Zydone, Zymase, Zyrtec, et cetera.

Now, Zyrtec is one of the names that I know has been reported to be confusing with Zyprexa. So there it is ranked number seven.

This other search on the right is based on the phoneme distance. So you convert the Zyprexa into this phonetic alphabet, and then you do in this case an edit distance search on it, and again Zyprexa is identical obviously, but the other ones phonologically that are similar to Zyprexa are Hiprex, Migrex, Zephrex, Zyprexa Zydis, and on and on down the list.

So you see how different measures produce different ranked lists.

The other thing that we've done is

integrated all of these other features. So you might want to know if you're a trademark attorney searching through this, well, Zyflo. I want to evaluate how similar this really is. So I click on Zyflo and I see, oh, it's a 600 milligram tablet, and I can click here and I say, oh, it's an oral route of administration, interpack size of 120. It's made by Abbott, and so on and so forth.

All of this data comes from the Multum Drug Lexicon, which is a free lexicon you can download off the Internet.

So simple illustrations of orthographic and phonological searches with some additional attribute information linked, but I think what's much more interesting is when we begin to search on multiple attributes.

So I've argued for a long time that you need to search on multiple attributes, not just the name, and that you also have to weight these attributes in some way.

So what I've done is on my laptop this

runs 1,000 unique products in this database. It takes about 90 seconds to run one of these attribute searches on my laptop at least.

I've taken Zyprexa again. I've entered its attributes, which it's a ten milligram tablet through the oral route. We have an integrated schedule because it's very to get schedule information. Each drug product actually has multiple schedules which depend on the age of the patient and so on and so forth. So we don't have schedule.

But I have assigned a weight of 60 percent to the name similarity, 15 percent to the strength, 15 percent to the dosage form, and ten percent to the route of administration, and here you find the results at the bottom. Not surprisingly Zyprexa in a ten milligram tablet is the most common product or the most similar product in the database.

And you go down and you see the most similar non-Zyprexa product is Zydone, which also is a ten milligram tablet, and Zyrtec, and you can click on these and find out all of their other attribute

information.

So we believe that these multi-attribute searches add a lot of value to the proposition of finding similar names in these databases, and that's just a simple demonstration of that point.

I have just a few more comments. I know I'm running towards the end of my time.

So one of the things we want to do is composite these similarity scores. I think Bonnie showed that each of these measures alone leave a lot to be desired, but you can take all of them together and then weight their combinations as they each contribute a little bit of unique information.

So here is an example we did in trying to predict expert judgments of similarity using these objective measures, and here's the actual results with an R squared of .4, which means 40 percent of the variance in expert judgments we can predict.

So on the horizontal axis we have objective similarity. On the vertical axis we have the predicted similarity based on that three or four

variable model. So you can, in fact, predict expert similarity judgments with some degree of accuracy using a combination of objective measures.

You can do the same thing with multiple attributes looking at similarity in the dosage form, similarity in the strength, the route, et cetera. You can a computer assisted pattern recognition to use? Yes. All of the stuff that I've described is computer assisted pattern recognition.

The general problem can be framed as a prediction problem in obvious ways with inputs and outputs, and you can tackle this through lots of different strategies, regression, discriminant analysis, and lots of different machine learning approaches.

There are problems with these methods. They're not perfect. They generate false -- just like with any search. Go to Google, the best search engine ever invented. Google will not give you perfect searches. There will be false positives, and there will be false negatives, things that it doesn't

retrieve.

The reliability of the data used for modeling is often suspect because it's based on voluntary reports. Like that USP list, there's no telling where those came from. Some are near misses. Some are real errors. We have no idea about the circumstances in many cases.

And that's about all. I think these measures ought to be used. In spite of their imperfections, they're much better than subject methods alone.

Thank you for your time.

(Applause.)

DR. GROSS: That was fascinating, Bruce. Thank you.

John Gosbee will speak next. The questions that he will address are:

How much weight should be placed on each of the review and data components, such as expert panels, focus groups, prescription drug studies, computer analysis or other issues in order to reach

an overall objective decision on the acceptability of a proprietary name?

And second, how much weight should be placed on the level and likelihood of patient harm?

DR. GOSBEE: Thanks very much.

Well, as some of the speakers said this morning and you hear from the introduction about me, I would say to some of you thanks for being around the party. I know you've been at this for a long time, like Mike, since the '70s and so forth, and I'd say to some of you welcome to the party.

And I think what has really been great about this conference is that one of the first times -- and I've been to 50 patient safety related conferences or meetings in the last four years -- that people have taken the time to understand at least some if not most of the complexities behind a seemingly simple problem, and so congratulations to everybody for doing that.

And as you'll see from my presentation, I think, there is a body of knowledge about the

complexity of why we make mistakes, slip up, confuse things that may potentially be as helpful as some of the discussion you've already heard.

Before I get into that, I just want to warn some of you, although when I do this sometimes I lose about ten percent of the audience, there's going to be some participation on your behalf, and so as an expert group, I'm going to ask a few questions.

The first one is going to be survey. How many people here when they use the restroom facilities went into the wrong bathroom? Anybody?

I see some smiles. Nobody is admitting it. Okay.

How many people with any of the doors that you've encountered since this morning, if you can remember, did you push the door instead of pull or pull the door instead of push? Anybody willing to admit that?

Okay. So a few more. And then the last question is: was that really a problem when you pushed instead of pulled or pulled instead of pushed?

And most of you -- we don't have time to go through everybody's answer -- would say it was an inconvenience or maybe the guy behind me slammed into me because he thought I was going to push right through, but you all probably survived that.

And I think you've heard a number of people say, especially most recently with Bruce that we are trying to look at error, but we're also looking at harm, and I think we can't break up the two pieces.

The other thing that struck me is that sometimes we do confusing things on purpose. Unfortunately this particular establishment burnt down in Alaska, but while it was there, it had a very interesting bathroom that was locate quite close to where the bar was where on purpose they put the handle next to the hinges, and of course, anybody new to the bar would go over there and push and pull with all of their might. Well, of course, all of the usuals in the bar would look and laugh at them.

And I think we've all seen magicians and

others that really play on the whole thing of confusion and really play into not just our sort of models of how things should work, but maybe our experience with things that we've just long since forgotten to take them for granted.

It's also interesting if you stay at this establishment, of course, that's now burnt down long enough to actually do sort of a "will that person remember the next time they go to the bathroom" and that happens to be correlated to how many beers they have in between their trips to the bathroom, and I do confess that I was probably one of those people who kept making that same mistake.

So what am I going to cover besides these sort of interesting stories? I'm going to go and emphasize that confusion goes well beyond naming of drugs, and I'm going to just cover briefly what failure modes and effects analysis and sort of a feeder to that, human factors engineering, and talk about how those two are related.

When I was given the task of sort of

describing not just FMEA, but how that fits into a bunch of other things, I sort of had the sense they were asking me to quickly describe the pros and cons of all the major religions in Western civilization. I mean, this is a huge task, and I'm going to try to boil it down to a few points and see if they'll stick with some of you.

I also want to, as a final sort of kick-off story, as I mentioned before, welcome to the party for some of you and thanks for being at the party for so long for others, but I read a very interesting book. I don't know if anybody is here from Upjohn or what used to be called Upjohn, but they had a really interesting history book where they encountered this look alike confusion back in the early 1900s, and I don't know if anybody knows this history, but evidently they made pills that did lots of stuff. You could ingest them and you could put them in glasses of water and they dissolved and you sterilized instruments in them.

And they kept getting these reports that,

you know, lo and behold, some people were taking these sterilization pills that were made out of mercury or whatever else.

And so they instituted an anti-confusion technique where they actually shaped the pills that weren't for ingestion into little, tiny coffins. That was a kind of interesting first effort at getting at this problem.

We've already covered a lot of this in the presentations. I will talk a little bit about sort of the ones at the bottom where much more eloquently you just heard from Bruce about the usual or expected delivery mechanism, but maybe then some other ones that haven't been covered as much, and that is sort of the metaphor or model that's conjured up as well as the appearance in cyberspace, and thanks to at least a few people this morning who identified that as sort of the next generation of issues when we think about how things are confused with each other.

So what kind of confusions do you see

here? For those in the back, I'm sure you can't read the labels on the things in the upper left-hand corner, but one is called EpiPen and EpiPen Jr., and EpiPen Jr. evokes what? Small, right? Kid-like, little one. And, in fact, for that particular metaphor model that's the real good thing to be evoking because that one is for I think it's children under X number of kilograms. EpiPen is for all the rest of us.

What does it look like? What does it look like out of its package? It's called EpiPen. For those who know how this EpiPen already works, don't raise your hand, but how many people think the EpiPen works by holding it and clicking the top? How many people think that?

How many people think it works by taking the cap off and the needle is there and then you go ahead and inject it?

Okay. A few more. How many think it works by stabbing?

Well, I hope you're not around when my

three year old is in need of her epinephrine because it actually works as sort of a stabbing or pushing. In fact, if you curl your thumb up over the top of this device, it won't work quite as well.

So this is not to pick on, you know, this particular company. We could go through a number of other examples. You've heard hundreds of them today where probably well intentioned naming evokes something or is confused with something where that wasn't the intended purpose when the person picked up on the name.

They asked me to talk about failure modes and effect analysis, and typically when someone does this, for those who don't know, you choose a topic or the area that you're going to look into. You form a team or an expert committee. Sometimes people can do this by themselves or multiple people do it independently, and you flow chart your process and your subprocess similar to some quality improvement techniques you've likely used or root cause analysis.

And then very systematically so that some

of the tools that were briefly mentioned earlier, sort of systematic guidance on how the expert or experts should follow through, you pick up on failure modes. You figure out maybe why those failure modes occur, and then you assign severity probability and visibility.

Layered on top of this, I would propose you could use and people have used the discipline of human factors engineering, and again, I don't have time to go into all of the aspects of human factors engineering, but we're going to do a few demonstrations to pull some of those ideas out.

But it's not just about designing systems. It also reveals a set of methods that look at needs and problems of the end user where you might encounter confusion or misunderstanding or getting lost and not knowing what to do with this particular system.

And then it also works on knowledge basis. What you actually heard recently for the last two speakers was really pulling up on knowledge bases

about what we know when humans interact with systems, using their senses, using their arms and legs, interacting with simple devices or organizations.

So some quizzes. This is your human factors engineering knowledge. Warning labels are effective in changing behavior all of the time if people are motivated for some people if labels are readable and understandable; some of the time if people are paying attention; not enough information to tell, and there's a citation there at the bottom.

Well, the answer is D. There's not enough information to tell. Most people would say -- some people would say one and others would say, well, this seemed to work pretty well sometimes, but the devil is in the detail, and if you look at the research, which is a huge body of research now, about just putting together what would appear to be straightforward things, like:

Don't drink this. You will die.

Don't put this on your arm. It may cause it to fall off.

You would think those are very straightforward sets of English words or direction and that people would follow them, but the literature is replete with people misunderstanding based on a lot of other factors, including the names of things.

If you want to move that dial to the right or towards the middle, do you rotate that knob clockwise or counterclockwise? How many people say clockwise?

How many people say counterclockwise?

People who say counterclockwise are probably thinking it's like turning things down or turning thins up, but you know, maybe it's like my faucet at home.

But most people came to this with a preconceived notion of "knobness." You already had an idea of what the knob should do in relation to the dial. It's not in our genetic code, but you've learned an awful lot in your lifetime, and you apply it in novel situations.

The same thing here. Which control knob

moves the dial on the right? Well, anybody with an old, beat-up 1970s or '60s version electric stove probably says it's random. You could pretty much turn any of those. Who knows which ones work, right?

(Laughter.)

DR. GOSBEE: Theoretically it should be the one on the right controls the one on the right, and we do that mapping or that association really without thinking about it.

And the purpose of these first three examples is really to tell you that, along with what we heard already about expert teams or expert groups is we're really carrying a huge amount of baggage about how we think things work, and when things don't quite work that way, we really are resistant against it and think, "No, no, they probably meant it to be this or they probably designed it to be that."

And we do an incredible job of sort of justifying to ourselves that something should work in a certain way or it should be that way as expected.

So another demonstration. Look at the

next slide and count the number of words in the paragraph that are repeated. Count the number of words in the paragraph that are repeated.

How many did you get? Three?

PARTICIPANT: Four.

DR. GOSBEE: Some people got five.

So you got those three, right? Everybody saw those three together. Count the number of the words in the paragraph that are repeated. More than that? Anybody got six? Maybe it's 14.

Everybody had a different interpretation of what the instructions meant. Some thought I meant repeated in a row. Some thought I meant repeated on the same line. Some people who have seen this demonstration before knew what the answer was.

But nevertheless, this is a very powerful phenomena that happens with people, and again, I'm glad Bruce and others have brought this up, where this idea of similarity or of matching or of confusion or of slip-ups or things going bad is a little more complex and ends up being very situation

dependent.

Now, that's going to make our job a lot tougher, and I know the hard hitting statisticians and others earlier this afternoon probably made your stomach tighten up as much as mine did in terms of what work is ahead of us.

But part of the reason that they do that is because of these other sort of conflicting variables, and I'll show you some data at the end where we've looked in the human factors methodologies and seen how well we do when we give experts a chance to identify bad stuff or poorly designed stuff.

So, Bonnie, you agreed to help out. I don't know if your colleagues in front of you can kind of move back a little bit. What I'm going to have you do is read the colors in the row as fast as you can, and so the top row, for instance, would be red, blue, green, yellow. So just read the rows as fast as you can, all three of them.

Can somebody turn her microphone on, please? There we go.

DR. DORR: Yellow, red, blue, red, blue, red, yellow, green.

DR. GOSBEE: Okay. Now what I want you to do is read the rows. So, for instance, the first row is going to be red, blue, green and yellow. So read the color of the words.

DR. DORR: Yellow, green, blue, red --

DR. GOSBEE: In the rows. So starting with Row 1 that's red. Then it's blue.

DR. DORR: Red, blue, green, yellow, yellow, green, blue, red, green, red, yellow, blue.

DR. GOSBEE: Okay, and read this one as fast as you can.

DR. DORR: You want me to read the colors or the words?

DR. GOSBEE: Colors of the words.

DR. DORR: The colors of the words. Green, yellow, red, blue --

(Laughter.)

DR. DORR: -- green, blue --

DR. GOSBEE: Okay.

DR. DORR: -- blue, red, yellow, yellow, blue, red --

DR. GOSBEE: I see Bruce laughing.

DR. DORR: -- green.

DR. GOSBEE: Bruce, do you want to take a shot at that? Come on.

(Laughter.)

DR. GOSBEE: Come on.

Now, she's got degrees in linguistics and computer science.

(Laughter.)

DR. GOSBEE: Certainly she can detach the part, the lobe of her brain that processes color from the lobe that processes words, right? I mean, this is simple stuff. We just have to be incredibly expert.

Don't we say this? We put instruction labels and warning labels that say, "Watch out. Ignore the color. Don't use the bad one, but just please ignore it."

Now, I wish this was completely funny. I

didn't have a medication example for you, but in the world of the medication delivery, which is compressed medical gasses, there are standards or people out there who have basically said, "Ignore the color. Just read the label."

People can't. In the real world, and I mean you have heard this from a number of people today, using more medication examples, the problem is you really can't.

And so when we talk about doing the studies, we talk about the expense of having to measure confusion or similarity. This is expensive, so to speak or relatively speaking, more so than just getting some people together and asking their opinion.

But I really do think in this case if you look at the cost effectiveness of it, it's way up there at least in my list.

Well, let's look at a little broader picture. I know this sort of wasn't in my charter, but I'm just going to spend a few minutes looking at

some things that a person did at Salt Lake City. He talked about sort of the analysis of confusion when somebody is retrieving medications from code cart drawers.

So he talked about how do you simulate this stuff to determine whether something is confusing or whether you can identify it, packaging, et cetera, and here's where he did do some real live user testing in a somewhat simulated fashion, and this was in his initial cart drawer, the laundry hamper approach: toss it in, hope for the best.

And you see the range there is between two minutes, 43 seconds and four minutes. Now, grant it that's a little bit contrived. You don't need ten medications all at once when you're doing a code, but it is a little worrisome that that number is up there.

So he went through many iterations and came up with a fifth version. He drove that number down to roughly one minute on average, and you'll note a few things. It actually lacks labels. He

used labels that actually drove the time up.

So as Bruce said before, sometimes the things that seem real obvious, if we just put like a bigger label, and I know that people talk about tall man lettering, like drawing attention to differences or similarities. Sometimes those things work; sometimes they don't.

The only down side is he said if you notice two arrows, the yellow and blue -- sorry -- yellow and red one -- I'm having problems with colors.

(Laughter.)

DR. GOSBEE: Yellow and red one there.

PARTICIPANTS: Green.

DR. GOSBEE: Green. Wow, I'm really having problems with colors.

(Laughter.)

DR. GOSBEE: You'll notice that he configured them such that you could read the label regardless of the orientation. That seemed to work out the best.

So why use human factors engineering? Well, human factors engineering tool like usability testing, which I'll show you some statistics on in a second, allows you to be more savvy in choosing the problem.

So, for instance, Bruce was saying, you know, not all of these things will be problematic in exactly the same way. So if you can understand these things in more of a context or usability testing, you can figure it out.

Add a human factors expert to your team and you can more accurately develop and test what you think the failure modes are and your solutions.

Everyone had to have a slide with data. Numbers are hard to process fast, but the second bullet down, expert evaluators, they've actually done studies of looking at confusion problems with software. Now, I recognize software is not the same as a drug name, but this is the data I have, and this is probably the only systematically gathered data that I know of in the human factors literature

looking at can't we just hire some experts and look at the software, you know, get the thumbs up or thumbs down versus do we really have to test it with end users and all that extra work.

And it turns out that the first check mark you see there, they did a study with five experts and they got 75 percent and with ten experts got 85 percent against the gold standard of 100 confusions or problems with, let's say, this piece of software.

But then with another piece of software and other experts, two experts found 90 percent of all the problems, but then they went to five other experts and they found 55 percent. Empirical data.

And then at the bottom they tried to -- same software, same types of problems -- they had an expert come in, and on average experts in that domain, let's say, the software for word processing, they found 20 percent of the problems.

Then they brought a human factors person in who didn't know anything about that particular

software. He or she found 40 percent, and when you had a combined human factors and expert, you rammed it up to 60.

So you can start to see looking at the variable that go into who's on the expert committee or the experts you have look at your particular system or device makes a big difference.

Now, the bottom, usability and user testing is more stable. That's where you put, and you've heard this many times, but not called usability testing, you put somebody in front of the system or device and have them use it in the way they would, and you can confound things by making them go faster or putting in other similar distractors.

And with four to six participants, you get around 90 percent. However -- and here's the really nasty one -- most of those usability studies when they look at the gold standard of what is confusing in medical devices, software, et cetera, the actual performance of people using those devices, their preference for features and whether they were

confused or not, how confident they were that they were not confused, so to speak, did not relate at all to their actual performance.

So people were saying, "Hey, this was a great XYZ machine," but when you looked at their performance, in fact, how many times they got diverted, how many times they got confused and had to start over again, that did not rate very well or correlate very well with what they said.

So I think some of the things you'll see from expert groups and expert opinions, you're going to need to have to watch out for this, and I second, third and fifth or how many people said you need multiple methodologies. You also need to understand where your methodologies fall short.

And I already mentioned this because I screwed up when I copied and pasted slides and forgot to cut it from the last slide.

And if you want to go to some place that's trying to do this, I know there's more than just this, but the University of Wisconsin and

Madison, VA, plus their College of Pharmacy have created medication safety and human factors courses. I think this is happening more in industry. Device companies have picked up on a lot of human factors engineering and so forth, and then the sister agency to the Center for Drugs, the Center for Devices, actually has a lot of stuff about the design and confusion and other issues related to good manufacturing practices with devices.

That's all I have.

(Applause.)

DR. GROSS: Well, thank you, John. That was fun and interesting. You caught us back up in time.

The last speaker, Dr. Bill Campbell, has three questions we've asked him to address.

The first is: what role should a premarketing commitment for a risk management place play in the approval of a proprietary name that has some potential for sound alike or look alike confusion with other marketed products?

Number two, what components of a risk management plan should be considered in order to minimize the risk associated with proprietary name confusion?

And last, what would be the measurable goal of such a risk management plan?

Bill.

DR. CAMPBELL: While he's getting my -- it's an amazing thing. You're all starting to look alike.

(Laughter.)

DR. CAMPBELL: And I suspect we're starting -- no, no. That's not me. And we're all probably starting to sound alike.

My presentation is the last on the docket and perhaps appropriately so because you might think of risk management as really the safety net for all the things we've talked about through the entire day. So in the context of what I'm going to be discussing, you might think of it as after all of the things, after all of the very sophisticated

techniques and preparations have taken place, what's the safety net in the marketplace that is appropriate, if anything is, in fact, appropriate.

And that, I think, is a good summary of the questions posed to me, and they were, as Peter said: what role should premarketing commitment for a risk management program play in the approval of a proprietary name? What components of a risk management plan should be considered? And what should be the measurable goals?

I think you already have heard and will certainly with my comments heard a set of repeating themes, and let me start by, first of all being clear what we mean by a risk management program.

The risk management program as I'm going to define it is a strategic safety program designed to decrease product risk by using one or more interventions or tools beyond the package insert, i.e., a safety net, and this comes from the FDA concept paper recently released and discussed at a hearing on risk management programs.

There are three categories, general categories, of risk management programs. There are a set of approaches having to do with specialized educational materials for health practitioners or patients; a set of approaches having to do with procedures or forms to increase compliance with approved or best practices reduced risk prescribing and use; and then a series of approaches having to do with modifying conventions prescribing, dispensing, and use of products.

What role should a premarketing commitment for risk management plan or program play? I would say follow three themes, and as Director and PI of the Center for Education Research in Therapeutics, I would be remiss in not telling you what our overriding theme has to do in the area of risk management programs, and that is to follow the credo of manage the risk and benefit the patient. I'll talk about that a bit later.

The second approach I'd like to bring before you is dealing with this what I consider

analogy of moving efficacy to effectiveness in the clinical trial, in the clinical arena in terms of efficacy and effectiveness in the drug naming arena.

And then lastly I wanted to take up a special question of can an approved risk management program reduce the time to market.

So first of all, manage the risk to benefit the patient. We've heard it already several times. There will always be risk. We cannot drive risk out of the system. It cannot be totally eliminated. What we should do is, in fact, welcome the opportunity to manage the risk because only through managing the risk can we deliver the benefit.

The challenge then is to identify the maximum acceptable risk, manage it and maximize the benefits.

Now, to the specific question, what component of a risk management program should be considered in order to minimize the risk associated with proprietary name confusion?

This is a set of components of risk

management programs that have been identified either in previously approved risk management programs in the literature and discussion and so forth, and as you can see, there are a rather large number. There are "Dear Doctor" or "Dear Provider," "Dear Pharmacist" letters. There's a whole area of active surveillance taking a specialize approach in selected emergency departments or ambulatory care clinics or health care systems to look at specific signals, passive surveillance, receiving signals as they come in, and trying to sort through them to identify the wheat from the chaff.

There are the sticker programs which you have seen both with Lotronex and with Accutane, and I think they're better referred to as attestation programs. They should be referred to, which means that someone in the system, the physician or the pharmacist or both, have attested to the fact that some decision or action has been completed. A pregnancy test has been taken, a diagnostic procedure has been performed, and then a sticker is added to a

prescription to attest that that, in fact, has taken place.

And we can move down the whole list of that. There are patient registration approaches, prescriber registration, restricted distribution, restricted prescribing, mandatory educational programs, a card system. I recently saw one where a person is given something that looks like a credit card, and that is the ticket into receiving the drug and also the risk management program. Eight hundred numbers, pharmacovigilant systems and so forth.

And that's not all of them. Educational programs in the form of journal ads, direct mailing, usual promotional activities. You may decide to credential a prescriber, not just register them, but credential them in the form of require them to complete an examination and pass at a particular score in order to prescribe.

Patient monitoring, pharmacist registration, and so on and so forth; a no refill policy; on and on.

And I think in some ways the most interesting things coming down the road are information technology solutions, such as computer physician order entry, Internet approaches, personalized electronic medical records.

Now, if you count up all of those back to the question of what should be the components of a risk management plan or which ones should be considered in order to minimize the risk associated with approving a proprietary name, we would find on the order of 20 that I've already listed.

Now, these are components that have been used or proposed in the approval of a drug to control for clinical risk and benefit ratio. But to be quite honest, the evidence is very, very sparse in terms of the effectiveness of any of these approaches either individually or in combination, and of course, when they're used, they're often used in combination.

To give you an order of magnitude of the problem, imagine testing each one of these individually, some 20, and the research agenda that

would be required to do that. Imagine then that each one of them are used in combinations from one to 20.

Imagine also, recognizing also that it's not just the combination, but it's the permutation that's also an effect in terms of measuring the effectiveness.

So you get a total potential number of risk management programs in the 20 zeros or beyond. So it's simply a huge problem, and we are just moving our feet into the water right now in terms of identifying what the components are. We have very little information about components of a risk management plan that should be considered not just in proprietary naming, but in measuring therapeutic risk and benefit.

So what I'd suggest is what we do know is that we also need to be creative and think of different approaches than just the ones that are on the list for current risk management programs, and thinking just a little bit creatively, we might imagine out of the box thinking of some different

components that we would use if we were to develop a risk management program to deal with proprietary name confusion.

You could come up with your list. I suspect a focus group or expert group could come up with an interesting list. Here are a couple I would suggest.

Perhaps written prescription only would be a part of a risk management program for dealing with proprietary name confusion. No verbal prescription allowed.

Perhaps attestation of the potential for the confusion, that is Tracleer, not Tricor. In other words, rather than a sticker, require the prescriber to attest to the fact that she identified not only the drug intended, but the drug that was potentially a confusion and alerted that as a rejected choice.

You could have a risk management program research design with actually multiple names and test the best and actually determine the best one in the

real world, and you could have actual -- in our information technology world, you could have instant prescriber validation by feedback, and we heard the NDC Health presentation today, which I think is moving down that road, where, in fact, it happens to be that it takes place at the pharmacy level, where a signal based upon potential look alike and dosage similarity is sent back to the pharmacy in real time saying, "Be sure that you don't mean this."

That's an attestation in real time using information technology. It's essentially what a sticker does, but a sticker doesn't do it in real time.

Now, if you move to physician order entry of prescribing, you could actually do the same thing at the prescribing level as well as at the dispensing level, and I think that's a fascinating approach that we ought to explore.

What we also need to recognize, of course, the problem with doing this is the problem of multiple false signals in the system that have

already burdened the system in terms of third party and prescription processing.

So we have to create and identify the balance between the signals that are useful versus the ones that are chaff.

The summary I would make on what are the components of risk management plan that we should consider are that there are no gold standards at the present time in terms of identifying components; only hypotheses to be tested.

And in terms of Susan Winckler's earlier commentation in the afternoon, she referred to the American Pharmaceutical Association and other organizations' desire to move away from a component or one-up approach in risk management program to a more systems approach.

I completely concur with that, and I think we really need to take that to heart.

Now, let me move to another part of the question, which I define the problem in terms of moving us from an efficacy into effectiveness. We

all understand the shift between efficacy and effectiveness in terms of therapeutic effect, but what we also need to think is what we're really talking about now is moving from efficacy data, in which case this comes from cognitive medical psychology application software as we've heard, focus groups, behavioral labs, case studies, qualitative methods, and so forth.

And we want to move that data into the real world of health care in prescribing and dispensing use of drugs. The problem is in the clinical efficacy and effectiveness world, we have the gold standard. We have the randomized clinical trial. So there is a basis for extrapolating information from research into the population.

In the problem we face today, we don't have that equivalent of the randomized clinical trial, and we've heard today a number of qualitative research methods that are at the front end, at the efficacy level, and inability to translate that into population, that is, effectiveness real world data.

Now, I want to be very clear as I say this because there is, I think, the impression from people in the qualitative research world, as I hear people such as myself and most of the people on this panel, in the quantitative world that there's an implied criticism. That's not true at all.

We recognize this is extremely rigorous research, qualitative research. It is just research that is not able to be translated into a population basis because of the lack of randomization in population representation.

So we have this very serious problem of the lack of a randomized clinical trial equivalent in moving us from efficacy to effectiveness.

Let me take a little bit of a side track on that now and let me talk a little bit about the role of the drug name in moving efficacy studies into effectiveness through the risk management program.

An efficacy study should be able to describe our expected risks, identify risks not previously suspected, provide an estimate of risk,

identify benefits not previously expected, confirm benefits, and so on and so forth.

Can an approved risk management program do all of those things and allow a drug to be moved more quickly to market by allowing us to accept perhaps a higher level of risk by being able to measure it in the marketplace?

Let me give you the hypothetical case of two different drug names that I've just made up, Appesate, which I would say is an appetite control, meaning appetite satiety or sating the appetite, or an existing drug name might be Apresolate, perhaps a high blood pressure controlling medication, a vasopressor of some sort.

So we could have two different names proposed, and under one scenario we could approve the requested name and move the drug into market due to a required risk management program.

Scenario B, we might defer approval of the premarketing study and not approve the drug and have no effect, that is, not reduce the time to

market.

Should, in fact, a risk management program be a mechanism for earlier approval of a drug in moving it to market is the question I would pose, and to answer the question, it gets back to the question that Dr. Lambert was raising from the floor and came up a number of times today.

What is the benefit of a drug name for us to make a decision that would allow us to accept and tolerate greater risk to identify additional benefit?

And to do that, we would have to identify a benefit in the name, simply from the name. Now, when can we do that? What would be the criteria for approving a drug contingent on a risk management plan?

Well, it's to short the time when no alternative therapy is available, when substantial therapeutic advantage exists for the new product, when therapy is for serious or life threatening condition, and when the risk and benefit can be very effectively communicates to all participants in the

system.

Now, move that to not the drug, but the drug name. Do any of these conditions apply then for allowing us to accept greater risk by moving a drug to market faster? Is it, in fact, the case that there's a situation where there is never an alternative name?

Well, I understand the terrain of approved names is something like 17,000. So there seems to be no shortage of creativity in finding names.

Is there ever a situation where there's a substantial therapeutic benefit for a new name? Well, would Viagra by any other name be more effective or less effective? It's a good question, and in fact, I'm open to the question to suggest that Viagra by itself have a therapeutic -- the name my have a therapeutic effect, and whether all of the other conditions apply or not, whether it's life threatening, serious, no available alternative therapy and so forth.

But I think we should be open to the question, and we should say that if, in fact, the name has a therapeutic benefit, it should allow us to accept greater risk, but the burden is on the sponsor to identify what that benefit is, which I think can be done by the research approaches discussed today.

Is there a treated condition that is made less serious by a name? That seems to me implausible, and the risk benefit of the name can be communicated. Well, I certainly think it can.

So I think the conclusion I would say that is in the general case of should we ever accept additional risk of any form in order to put a confusing name on the market, I would say the case is unproven, but I would also say that we should be open to proving the case, and in fact, that might be possible at some point in the future.

What are the components of a risk management plan that have been shown effective in minimizing risk associated with proprietary name confusions? There aren't any. I think that's an

easy one.

Now, I would say there is face validity to at least two of the components, and that is restricted distribution and restricted prescribing. While there would not be data to suggest this, I do believe it's logical on the face that if you restrict prescribing and dispensing to a certain category and have the stick of eliminating that from a practitioner's armamentarium, that's a very powerful lever and on the face of it you could, I think, argue that those are two effective components at the beginning.

I would say that unproven would define all the other of those 20 or more components that I've identified. So I think we would be left with a couple of hypotheses. One is the effectiveness of individual elements of risk management programs is not know, to answer the question, and the effectiveness of any combination or permutation of those is not known, and thirdly, that's a very large and unfortunately unfunded research agenda.

What should be the measurable goals of a risk management program, if in fact we decide that that is the proper approach? I think there are four questions that need to be answered.

First of all, what's the base line of error? What's the minimal acceptable risk or error? What is our measure of success in a risk management program? And then what's the target?

Let me take those in turn. I think we, first of all, need to determine what is the baseline in order to develop an effective risk management program, and the baseline I would argue is that error rate for a proprietary name with no projected look alike/sound alike confusion. I don't know what that baseline is. Let's call it alpha, but we do know it's greater than zero.

It requires us to have some knowledge of risk, to have some knowledge of current practice, prescribing, dispensing, and use.

What's the maximum acceptable risk? Well, again, we don't know that number, but I think

we could define it as the acceptable error rate for a proprietary name with potential look alike/sound alike comparators.

That is, if the baseline is the irreducible minimum, that is, as Dr. Laser was saying, the vortex, when a drug is thrown in the vortex without any confusing comparator, that's the baseline.

When the drug is thrown into a vortex with a look alike/sound alike distractor and we accept that, that's the maximum allowable list. I call that beta, and I think we could say that beta is greater than alpha, alpha is greater than zero, and what we might know from the Barker and Flynn study is that it might be on the order of .13 percent as a starting point for discussion.

But this requires not only information from the baseline data, but it also requires us to have knowledge of benefit and risk of the proposed name, as well sa knowledge of risk and benefit of all distractor or comparator names.

Next we need to know for a risk management plan or risk management program what's our measure of success. I think success in a risk management program has to be defined as a range of error or range of risk that is equal to or less than the maximum acceptable risk, beta, but equal to and greater than a baseline risk, alpha.

In other words, our measure of success equals gamma where that is someplace greater than alpha and less than beta.

What are the targets of a risk management program? A target -- and this is the critical question -- a target is a specific quantitative goal for the error rate established a priori by a risk management program, i.e., it is an expected rate, and it is a point at some point between the area of what we define success or the gamma areas, that is, some place between that range of rates equal to or less than the maximum acceptable risk, but equal to or greater than the baseline risk, and any point on that can be defined as a target for our risk management

program and has to be defined in advance for us to be able to move to a risk management program.

So we have a couple of options. We don't have to keep naming things the way we do. We could take the approach of hurricanes and tropical storms. We could have gender specific names alternating between name and female between particular storm, and the name would acquire the attributes of the drug.

Floyd happened to be a very, very powerful storm, but that was because the storm was powerful, not because there was anything intrinsically beneficial or risky about the name Floyd.

Thoroughbred horses do the same thing. Initially they're just given an alphanumeric designator. Was Secretariat faster because he was named Secretariat? No.

But the Option B is to continue our status quo, and that's surely what we will do, which is to continue expanding the terrain of our existing names, perhaps exponentially, and use first come,

first name decision making. We will class name drugs by competition as Bruce identified the prefixes and suffixed being named on the basis of competitive submissions, and drug will acquire the attributes of the name potentially.

And I raise the question again. Is sildenafide more effective because it has the name Viagra than it would be if it had a different name?

And I think that's a very interesting question and a conundrum that really gets back to the fundamental question of: is there a benefit to any proprietary name that can be measured? And if there is a benefit, can it move us to then accept some balancing?

So in conclusion, let me say risk management programs can improve our risk benefit ratio, but the choice of individual elements are an optimum combination requires a huge amount of primary research that has not yet been conducted.

In order to have an effective risk management program, there must be measurable

quantitative goals for baseline risk, acceptable risk, success, and targets, and those must be all identified a priori from quantitative research methods that do not currently represent the state of the art.

And, thirdly, given the state of the art or research and proprietary name related risk management programs, this is not a mechanism for reducing time to market or accepting risk in any other form.

Thank you.

(Applause.)

DR. GROSS: Okay. A great session. We now have some time for questions for about ten, 15 minutes before we'll do the sum-up. Does anyone have any questions?

Yes.

MR. KOLLURU: Rao Kolluru of Bioxy Source.

Can you hear me?

PARTICIPANT: No, we cannot.

DR. GROSS: Flip it on. Yeah, go ahead.

MR. KOLLURU: Okay. John, you referred to --

DR. GROSS: Hold it closer to you.

MR KOLLURU: You referred to the failure mode analysis. I was wondering how far downstream you saw those failures. In other words, much of what I heard today seems to be stopping at the first effect of an error or a fault, but the subsequent notes may be more serious.

To give you an example, let's say that a contamination is detected in public water supply, and the public health officials decide to cut off the water supply, but the same water may be used by the fire department, and the risk, of cutting off the water supply may be higher there than whatever risk is posed by drinking the water.

I just wondered, you know, how far down do you carry that. Typically in engineering we look at a number of different nodes, maybe three or four or five levels beyond the first effect.

So in other words, you know, what is the propagation effect? Don't we need to look at those propagation and dependency relationships?

DR. GOSBEE: In general, yes. When it comes to the naming, proprietary naming, for today I think that was very well explained by a number of speakers in terms of, you know, can we say that, you know, if a name has some confusion risk, you know, is there any sort of benefit to then going ahead and approving that particular name to get it to market faster because it will have an unintended consequence of it taking longer and things like that?

I don't know enough about the process to say if that happens. For failure modes and effects analysis in general, one of the deficiencies of that particular approach, as well as root cause analysis in any of them, they're only as good as the people who have knowledge about what bad things can happen and what conceivably are bad things that happen if you fix the first bad things that happen.

And that's why I was very happy today, as

I mentioned that the complexity of the problem has been very well outlined today. I think this has been a great conference to show us that you really do need to carry these evaluations out in a sophisticated way and have a depth of understanding, and then you make the best decision.

DR. GROSS: Okay. Any other questions?

(No response.)

DR. GROSS: What a bright audience. You understood everything that wa said. If that's the case, shall we start --

DR. LAMBERT: I think there will be a brief quiz now if there are no questions.

(Laughter.)

DR. GROSS: If that's the case, I'd like to thank the panel for their contributions. They're wonderful.

And, Michael, shall we start the sum-up a few minutes sooner?

(Applause.)

MR. LEE: I think each of the panel

moderators is going to try to sum up their sessions for the day. So I'm going to start and Jerry Phillips will finish wrapping up.

I'll have to put my glasses on because I made notes, and it'll be difficult reading my own notes.

My hope for this particular conference was that putting a spotlight on the methods that are currently being used would help us to improve the methods dramatically, and I think that it passed my expectations tremendously.

I think my gratitude goes out to the experts who came today and shared with us their best thinking on the subject because I've certainly learned a lot. I had some fairly strong preconceived notions about where we ought to be headed based on some years of experience wrestling with the problem, and I've got to rethink some of those things because I think we learned an awful lot today.

Let me just point out just a few things. At our 12:45 p.m. panel, we highlighted a few

things. For example, difficult to really do a quantitative analysis until we get some things in order, and we probably need to start with a qualitative evaluation of trademarks first.

And from Shari Diamond, I think, in her session, there's probably a lot that can be done to improve the process in terms of the way a questionnaire might be designed.

Also, from Kaz's presentation, handwriting has always been so nettlesome to try to deal with. It's frustrating at times. You wonder whether or not any of the data is really relevant or reliable, and yet Kaz's technical system, there might actually be some assistance from the technology in that area that could help us.

I thought the "groupthink" slide was just a tremendous slide to show how to try to keep the expert panels from getting to biased and to keep them honest, and I thought that was just an interesting slide about how an expert panel should operate.

And also the notion that we're probably

headed toward the computer mediated decision help, mixed in with expert panels and some of the technology that's coming along, is probably going to give us better decision making results.

In the afternoon phonetic similarity is a difficult thing to evaluate, and yet we saw some perhaps improvement in that ALINE research that was done, and Dr. Lambert always gives us things to think about. Frequency may be as important or more important than a certain level of similarity in some respects.

Also, I thought that the work on the database that Bruce is doing and providing a lot more information in a single database where you can pull those factors up immediately when you're looking at the trademarks helps to evaluate them, although weighting them becomes rather a subjective measure right there.

I thought the later discussions -- one of the things I wrote down was "discordant cues cannot be ignored." That was a very effective presentation,

and I think we have to keep that in mind in the brand name and in the total packaging when we put things together, maybe even the logo design of the brand name.

Also, add human factors engineering expert to your team, something else I wrote down.

In the final discussion, the one reaction I had there was on number one, no one would ask for a risk management program to expedite bringing a confusing name to the marketplace because we don't want to bring confusing names to the marketplace.

Also, I think the whole day was about recognizing that right now we don't have a good handle on how to evaluate whether brand names are causing the problem or not. We see a lot of numbers. We haven't been able to evaluate them to conclude just how difficult the problem is.

And the one thing I would put out there is if you're going to put a name for a product, you certainly wouldn't do it randomly. You do it with a certain amount of forethought with an effort to make

your names different from all the other names in the marketplace, and that's what trademark attorneys do, and now assisted with many of the experts in the medication area, I think we're doing a better job.

Dr. Lambert published a paper in which I think he showed an analysis of the neighborhood distances, if you will, of trademarks that are in the trademark registry, and basically we're doing a pretty good job based on those kinds of objective measures.

So I think we ought to keep in mind that the proprietary names that are out there, many, many of them are done with an effort to keep them different. That doesn't mean you're always successful, but the effort is put in first before the name goes into the market place.

DR. COHEN: Thanks, Bob.

Well, I'll try to do my best in summarizing the session that I moderated.

First of all, I have to say I agree with Dr. Gosbee, what he said before when he first started

talking about, you know, going to a lot of different conferences, and it's pretty rare to actually address a specific subject and not just hear what's wrong, but also make some recommendations and take some recommendations home for changes that might actually help the situation.

We certainly hear that today, and I hope that, you know, there will be a future meeting or other times when we can get together and discuss some of these and try to build a system that could eventually be tested as a gold standard.

I think we heard from Dr. Strom that we had the wrong question in mind when we asked him what was an appropriate sample size. And that's a big deal, I think, for all of us at FDA and the testing companies and pharmaceutical companies that are sponsors of these new products. No one really knows. There's nothing wrong with 200 people. There's nothing wrong with 30 at this point. We really don't know what the appropriate number of individuals is in a sample size for testing.

And, therefore, we can't make any judgment calls right now, and at least for now we should go on doing what we're doing until we look at it in a different way, and I think what Dr. Strom was saying was what we're doing now is qualitative. Obviously this is important work. I think everyone recognizes there are some names that might have reached the market and potentially cause problems.

I've been stopped before that actually happened. So it is qualitative and you can't come up with a sample size based on the information that he gave us about the factors that are needed to make that calculation.

What is needed is to come up with a standard, take information that we learn today, add to it, and then do the appropriate type of testing, different types of testing. He gave us some ideas of the kinds of things that we could do to do appropriate testing, not of a specific drug name, but of the technique, the method that is actually used. So I think that was an important contribution today.

He also said that probably the best way to do this study or the testing was direct measurement of an error rate, and that would require some type of simulation, and I think that brings some complexity to the testing process. It's very difficult to do that and then also take into account all of the latent failures in this systems. There's little errors out there in everybody's system that contribute that we saw when Tim Lesar this morning was talking about the vortex and then listed all of these other factors that contribute to it.

That's not something you can do in a simulated environment. It makes this job extremely difficult to do accurately. So I think that was an important contribution that we heard today.

For Shari's talk what I heard was several different things. She talked about not being able to know how good the experts are in predicting what might go wrong, that is, what people in the real world might find that might go wrong. So it's crucial to conduct some tests to evaluate the degree

of she called it that correspondence before we can rely on expert predictors. So that is something that we have to take into account.

She also told us that she gave us some ideas for requirements of expert panel members. She told us that using the Internet, it's certainly possible to do self-administration of these questionnaires that are being used by various organizations, that that's feasible with Internet access, and that might be important because perhaps we may need a larger sample size, and that would facilitate that process.

She told us that one of the things she would like to see is some time limited exposure to the graphics that are used now for the drug names, the handwritten samples, and that's not to my knowledge being done to any large extent at this point. So that might be something that we need to think about in building this potential gold standard that can eventually be studied if, hopefully, the appropriate funding can be found.

She also talked about open ended questionnaires and when they may be valuable and when they might not be valuable. She said that they should not be used when testing for comprehension or recall. Basically that generally doesn't take place with safety testing. That's not what happens here.

With safety testing you are given a list of conditions that you would appropriately use open ended questionnaires for.

She also talked about focus groups and said that by themselves they are good for generating ideas, but very weak for evaluating individual reactions to the specific stimuli, the names that are presented and the other information presented. So that was also a very important contribution.

I think the one thing that took me by surprise was Kaz's presentation. That was the first time I ever saw the availability of that software, the handwriting recognition software that actually could be used, provided that we had appropriate graphics, some type of a system to detect perhaps

that we have in the system other drug names that might look similar. And there was very good information that we got from that.

His software, and I assume that there might be other companies that manufacture software like that, segments cursive handwriting and compares these segments to graphic representations.

One type of testing that could be done is without a large database of graphic representations of existing names which, as we know, according to Kaz does not exist at this point and would have to be built at some expense; that you could at least in the interim do a graphic representation of the name and then test it against databases that currently exist that are not graphic, and at least that would have some advantages for doing that.

That this system even worked with sloppy handwriting and that it didn't require training. And I think by training the implication for me at least was training of the software, not training of individuals on how to, you know, run the operation.

I think that's what that meant.

And that it could even check when there was a misspelling for how similar the misspelling might be to something that's already in the database.

I think we also heard, based on some of the questions, that it could do in addition to just the graphic comparisons of drug names, you could also include other information similar to what Dr. Lambert was talking about when you're talking about evaluating a drug product and not just the drug name.

So I think that was a very valuable session, and I think we learned a lot from it that could be included in future plans for improving the testing methods.

DR. GROSS: Thank you.

Well, thank you for including me in this meeting. I learned a heck of a lot.

I'd like to compliment Michael Cohen, the FDA, and PhRMA for conceiving having this meeting in the first place. I can tell you every meeting I've been at with Michael Cohen, whether it's at the FDA

or the Joint Commission he always brings up look alike and sound alike names and the problems with it, and to his credit, he has kept this high on the agenda for us to try to improve health care. So my compliments.

Our session was on decision analysis tools. it was really on the interaction between humans and machines

We first heard from Rick Shangraw who talked about the value of expert committees, and in my world of clinical practice guidelines, expert committees are the lowest level of evidence after randomized controlled trials and other controlled studies.

However, there are times when we don't have that information available and even when we do we still need expert committees to put it all together.

So a very important presentation. Clearly those efforts are better than individual decisions being made by some guru.

He mentioned the appropriate panel size and the credentials of participants and warned against people who are too into the particular field because they may intimidate or dominate the discussion, and the importance of diversity in putting together a group.

How to meet and be the most effective? I don't know how many of you -- well, I'm sure all of you participate in conference calls. I'm not sure how effective those things are. I usually spend half the time doing my mail while I'm participating in the conference calls.

So I think it's important to hear what are the ingredient that will make for a successful effort, and you certainly have to go beyond simple faceless conference call. So I think those were very important to consider.

Dr. Bonnie Dorr was fascinating information looking at the orthographic and phonological assessment. I personally had used Soundex in a computer program to identify patients

many years ago, and it's fascinating to see the progress that has been made sine then.

With her colleague, Dr. Kondrak, it was interesting that the ALINE, the acronym ALINE, turned out to be the best approach. I think the thing that's going to come up over and over again -- I know it's also true in trying to implement practice guidelines -- you have to use a multifaceted approach here to determine what the best and least confusing names are. We're clearly going to have to use multiple methods.

With 399 names out there that can be confused, we do have a problem.

Dr. Bruce Lambert brought to our attention the fact that it's not just what's in the name. It's what the whole drug is all about and pointed out that all medical errors do not cause harm, and again, we need to look at the bigger picture.

I thought that his software demonstration was fascinating. I'd love to see more of that.

Again, putting together a composite score using multiple measures and a regression model will be the best way to go.

Dr. Gosbee warned us what to watch out for the next time we go into a bar.

(Laughter.)

DR. GROSS: I thought that was very helpful. He pointed out with great illustrations the things that are conjured up in our mind when we see certain names or certain physical objects, how we may assume more than we should.

I know when my kids put something together the last thing they look at are the instructions, but I think certainly in the field we're in being aware of human factor engineering is very important and made us realize that people with knowledge in this area should be part of the teams when we try to decide on what names will or will not work for a new drug.

Dr. Campbell finalized our session and talked about risk management programs, and that was

very useful. He pointed out that there will always be risk, and we just have to learn how to manage it. He summarized the components of risk management programs. I thought that was very useful. "Dear Provider" letters, active or passive surveillance, attestation efforts, patient registration, and many others were shown on his slide.

The ones that hold the most promise probably because we know the least about it are the information technology methods, computerized provider order entry and electronic medical records.

He had a number of suggestions for components of risk management programs for names, and I think all of that information was very useful.

It was sobering to realize that there are no gold standards, and moving from the purity of a randomized controlled trial, namely, efficacy, to the real world of effectiveness will be a challenge, and it will be even more difficult because we don't have randomized controlled trials that we'll be able to refer to.

I think we should look at and test his quantitative methods for deciding how to determine what the maximal acceptable risk is, what measures of success are, and determining targets of a risk management program.

So I thought it was a wonderful program. It was a wonderful day, and thank you very much.

(Applause.)

CAPT. PHILLIPS: In conclusion, from FDA's perspective, this has been a wonderful meeting. It has been a good dialogue. It has been a first in a discussion that we will continue and having probably another public meeting to discuss in the future. So this is just the beginning of a dialogue.

I would like to on behalf of PhRMA, ISMP, and FDA thank the speakers and everyone for being here and for your participation today.

Thank you very much.

(Whereupon, at 5:00 p.m., the meeting was concluded.)