FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

MEETING OF THE

DRUG SAFETY AND RISK MANAGEMENT ADVISORY COMMITTEE

8:05 a.m.

Thursday, December 4, 2003

Holiday Inn
Two Montgomery Village Avenue
Gaithersburg, Maryland

ATTENDEES

COMMITTEE MEMBERS:

PETER A. GROSS, M.D., Chair
Chairman, Department of Internal Medicine
Hackensack University Medical Center
30 Prospect Avenue
Hackensack, New Jersey 07601

SHALINI JAIN, PA-C, M.B.A.
Executive Secretary
Advisors and Consultants Staff (HFD-21)
Center for Drug Evaluation and Research
Food and Drug Administration
5600 Fishers Lane
Rockville, Maryland 20857

MICHAEL R. COHEN, R.PH., M.S., D.SC.
Institute for Safe Medication Practices
1800 Byberry Road, Suite 810
Huntington Valley, Pennsylvania 19006

STEPHANIE CRAWFORD, PH.D., M.P.H.
College of Pharmacy
University of Illinois at Chicago
833 South Wood Street, M/C 871
Room 258
Chicago, Illinois 60612

RUTH S. DAY, PH.D.
Duke University
Department of Psychology: SHS
Flowers Drive, Building 9, Room 229
Durham, North Carolina 27708

CURT D. FURBERG, M.D., PH.D.
Department of Public Health Sciences
Wake Forest University
Medical Center Boulevard, MRI Building
Winston-Salem, North Carolina 27157

JACQUELINE S. GARDNER, PH.D., M.P.H.
Associate Professor
Department of Pharmacy
University of Washington
Health Sciences Building, Room H-375
1959 Pacific Avenue, N.E.
Seattle, Washington 98195

ATTENDEES (Continued)

COMMITTEE MEMBERS: (Continued)

ERIC HOLBOE, M.D.
Yale University School of Medicine
333 Cedar Street, 1074 LMP
New Haven, Connecticut 06510

ARTHUR A. LEVIN, M.P.H., Consumer Representative
Director
Center for Medical Consumers
130 McDougal Street
New York, New York 10012

LOUIS A. MORRIS, PH.D.
Louis A. Morris & Associates
8 Norman Court
Dix Hills, New York 11746

ROBYN S. SHAPIRO, J.D.
Ursula von der Ruhr Professor of Bioethics
Medical College of Wisconsin
Center for Study of Bioethics
8701 Watertown Plank Road
P.O. Box 26509
Milwaukee, Wisconsin 53226

BRIAN L. STROM, M.D., M.P.H.
Professor
Department of Biostatistics and Epidemiology
425 Guardian Drive
Blockley Hall, Room 824
Philadelphia, Pennsylvania 19104

SPECIAL GOVERNMENT EMPLOYEES: (Voting)

JEFF BLOOM, Patient Representative
Washington, D.C.

ATTENDEES (Continued)

GUEST SPEAKERS: (Non-voting)

BONNIE DORR, PH.D.
Associate Professor
Department of Computer Science
University of Maryland

SEAN HENNESSY, PHARM.D., PH.D.
Assistant Professor
Department of Epidemiology and Pharmacology
Center for Clinical Epidemiology and Biostatistics
University of Pennsylvania School of Medicine

MIRIAM BAR-DIN KIMEL, PH.D.
Senior Project Manager
MEDTAP International

ROBERT E. LEE, JR., J.D.
Assistant General Patent Counsel
Eli Lilly and Company
Representing Pharmaceutical Research Manufacturers of America (PhRMA)

KRAIG SCHELL, PH.D.
Assistant Professor
Department of Psychology
Angelo State University

RICHARD F. SHANGRAW, JR., PH.D.
CEO
Project Performance Corporation

FOOD AND DRUG ADMINISTRATION STAFF:

JERRY PHILLIPS, R.PH.
PAUL SELIGMAN, M.D., M.P.H.

ATTENDEES (Continued)

ALSO PRESENT:

DOUGLAS BIERER, PH.D.
SUZANNE COFFMAN, PHARM.D.
CLEMENT GALLUCCIO
BRUCE LAMBERT, PH.D.
PATRICIA STAUB, J.D., R.PH.
MAURY TEPPER III, J.D.

C O N T E N T S

Issue: Current Screening Methods to Assess Sound-alike and Look-alike Proprietary Drug Names in order to Reduce the Incidence of Medication Errors Resulting from Look-alike and Sound-alike Names

* * *

C O N T E N T S (Continued)

P R O C E E D I N G S

(8:05 a.m.)

DR. GROSS: Good morning, everybody. I'd like to start the meeting. If you plan on going home today, we should start the meeting now.

I am the chair of the Drug Safety and Risk Management Advisory Committee. My name is Peter Gross. I'm the Chair of the Department of Medicine, Hackensack University Medical Center.

We have a very interesting agenda today.

I'd like to go around and introduce the members of our advisory committee or have them introduce themselves. We will start with Brian Strom at my left.

DR. STROM: I'm Brian Strom from the University of Pennsylvania School of Medicine.

DR. CRAWFORD: Good morning. Stephanie Crawford, University of Illinois, Chicago, College of Pharmacy.

DR. HOLMBOE: Eric Holmboe from Yale University.

DR. LEVIN: Arthur Levin, Center for Medical Consumers.

DR. MORRIS: Lou Morris, Louis A. Morris and Associates.

MR. BLOOM: I'm Jeff Bloom from Washington,

D.C. I'm an AIDS patient advocate in Washington, D.C.

DR. DAY: Ruth Day, Duke University.

DR. COHEN: Mike Cohen, Institute for Safe Medication Practices.

DR. GARDNER: Jacqueline Gardner, University of Washington, School of Pharmacy.

DR. FURBERG: Curt Furberg, Wake Forest University.

MS. SHAPIRO: Robyn Shapiro, Center for the Study of Bioethics, Medical College of Wisconsin.

MS. JAIN: Shalini Jain, Executive Secretary for the advisory committee, representing the FDA.

DR. GROSS: The two people from the FDA that are at our table are Dr. Paul Seligman, who is Director of the Office of Pharmacoepidemiology and Statistical Science, and Acting Director of the Office of Drug Safety, and to his left is Jerry Phillips, Associate Director of Medication Error Prevention at the FDA.

Shalini Jain now will go over the conflict of interest statement.

MS. JAIN: Good morning, everyone, and thanks for attending our meeting today.

The following announcement addresses the issue of conflict of interest with respect to this meeting and is made a part of the record to preclude even the appearance

of such at this meeting.

The topic of today's meeting is an issue of broad applicability. Unlike issues before a committee in which a particular product is discussed, issues of broader applicability involve many industrial sponsors and academic institutions.

All special government employees have been screened for their financial interests as they may apply to the general topic at hand. Because they have reported interests in pharmaceutical companies, the Food and Drug Administration has granted general matters waivers of broad applicability to the following SGEs, or special government employees, which permits them to participate in today's discussion: Dr. Michael R. Cohen, Dr. Ruth S. Day, Dr. Curt D. Furberg, Dr. Peter A. Gross, Dr. Louis A. Morris, Dr. Brian L. Strom.

A copy of the waiver statements may be obtained by submitting a written request to the agency's Freedom of Information Office, room 12A-30 of the Parklawn Building.

Because general topics could involve so many firms and institutions, it is not prudent to recite all potential conflicts of interest, but because of the general nature of today's discussions, these potential conflicts are mitigated.

In the event that the discussions involve any

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

With respect to all other participants, we ask in the interest of fairness that they address any current or previous financial involvement with any firm whose product they may wish to comment upon.

Thank you.

DR. GROSS: For the record, I'll read the main issue being discussed today. Current screening methods to assess sound-alike and look-alike proprietary drug names in order to reduce the incidence of medication errors resulting from look-alike and sound-alike names.

Now I'd like to reintroduce you to Dr. Paul Seligman, Director of the Office of Pharmacoepidemiology and Statistical Science and Acting Director of the Office of Drug Safety.

DR. SELIGMAN: Good morning. It's a pleasure this morning to welcome back our Drug Safety and Risk Management Advisory Committee, those of you who are going to be making presentations this morning, as well as all of you who will be participating in today's discussion. Today we have a full committee assembled, and I thank you all for your time and effort and consideration in being here today.

Peter has introduced the topic up for discussion for today which is to look at current screening methods to assess similarities amongst proprietary drug names. As some of you may realize, this topic was scheduled for discussion on September 19th of this year. This discussion seems to bring along the weather. Unfortunately, the meeting was canceled because Hurricane Isabel came roaring through and forced the last-minute cancellation, and I apologize to those of you who either en route or actually had arrived here in Washington just prior to that last-minute cancellation.

At today's session we're going to be hearing from several speakers who will elaborate on a number of different drug screening methods. I'm looking forward to exploring this issue with the help of Dr. Gross and the other advisory committee members, as well as our guest speakers. There are a number of questions that we will formally pose to the committee for consideration which will be presented at the end of these presentations and prior to this afternoon's discussion. So once again, I'd like to take this opportunity to welcome everyone again and thank our committee.

With that, I think I will start the program this morning by teeing up the first topic, which is advancing the science of screening proprietary drug name

review.

The underlying basis for our discussion today is that there are a substantial number of medication errors that result from confusions caused by look-alike and sound-alike names and confusing packaging and drug labeling.

In the 1999 report from the Institute of Medicine, To Err is Human, the IOM report proposed that the FDA require drug companies to test proposed drug names for confusion.

In November of 2002, the Department of Health and Human Services Committee on Regulatory Reform called for the FDA to shift the responsibility for conducting this kind of review and testing to the industry.

In June of this year, in cooperation with PhRMA and the Institute for Safe Medication Practices, we held a well-attended and interesting public discussion here in Washington, which was really the first attempt to explore the current methods to screen proprietary drug names for similarities. It was an outstanding, interesting, engaging, and robust discussion, and basically what we heard was that the current approach, which is largely qualitative, isn't consistent, nor can most approaches at present be validated or reproduced. I'm going to talk in greater detail about more of the comments that we heard in that meeting, but that was sort of the overall message that

we got out of that discussion.

There is a variety of approaches that can and have been used to screen drugs for proprietary names. You're going to be hearing experts this morning sort of delve into those particular topics. I'm going to take a moment this morning to sort of talk about some of those and some of the concerns and issues raised by these particular methods.

The first method is the use of basically expert committees, people knowledgeable in pharmacy, people knowledgeable in issues related to behavioral sciences, et cetera. Basically in the area of expert committees, which is essentially assembling groups of 8 to 12 participants to look at names, I think one of concerns that we have is that there's not much research in these areas. If experts panels are to be successful, they need to be run consistently to be useful. There has to be an establishment or clear understanding of what the baseline level of expertise that is needed for these expert committees. And as always, whenever you assemble groups of people together to review things, there is a tendency for group thinking, if you will.

There is a whole host of challenges related to surveys and questionnaire designs, including how to design surveys in anticipation of marketing a product prior to

that product actually being available, limits on experts' ability to predict errors, the need to consider how one might develop simulated circumstances that accurately reflect a pharmacy or prescribing environment, and what ways one might consider the use of focus groups in generating ideas, although clearly these are approaches that are, by their nature, weak in evaluating individual reactions to stimuli.

The engineering world uses a variety of approaches in failure mode and effect analysis that range from picking expert committees and teams to detailing of flow charting processes to determine root cause analyses of errors, to using using tools that systematically go through each step to determine essentially what's not working and why it's not working, and to assign a level of severity, as well as visibility, for a particular problem. The degree to which these kinds of techniques can be applied to evaluating and assessing proprietary names has yet to be tested, but I think there are many lessons to be learned from the world of failure mode and effect analyses.

There is a variety of handwriting recognition techniques that combine certain basic elements of handwriting that are similar to all handwriting techniques that involve pattern recognition of writing a proposed name and developing databases of graphic patterns for all

existing drug names to make a comparison. So we'll hear about some of these today as well.

There are also computational linguistic techniques that can be applied. This is an area that we at the FDA have been particularly interested and have worked closely with a contractor to develop a system which allows us to systematically screen the names using a software algorithm that allows us to look at phonetic strings and groups of letters and to do essentially orthographic and phonological matching and screening of names.

It's also possible to consider standard study design and sampling techniques. You'll hear a little bit this morning of the approach that we use at the FDA to essentially conduct our own internal sampling of names. Although this is the approach that we use and I think we've used it with some degree of success, there clearly needs to be some standardization of this approach, tests for reliability and reproducibility and validity since the work that we do at the FDA, while valuable, does not have a gold standard against which we can measure the results of our work.

As I indicated, there is also a variety of computer-assisted decision-based analyses that can be a powerful driver in terms of looking at prescribing frequency, looking at potential harm that certain name

confusions can cause, as well as developing objective measures to demonstrate reliability and predict the probability of human error.

Another key issue for us in this era of risk management is what role risk management programs play. Are there situations where certain name confusions, because of the potential risks of the drugs, may be more acceptable than in other situations where a potential name confusion can be devastating or life-threatening?

Clearly in an era where we are looking at all elements of managing risk and how to validate and understand how these elements and tools function and how well these plans work, we're clearly interested in knowing as well whether risks associated with names and naming can also be managed in the post-marketing environment and whether one could design risk management plans around limiting errors associated with potential confusions of names. Many of the elements in our upcoming risk management guidance talk about the need to demonstrate baselines of error, demonstrating goals for programs and measuring the success of these programs. Can these techniques and principles be applied as well to errors and problems caused by name confusion?

So basically at the public hearing last June, we heard I think the following major themes.

First, the need to adopt a more systematic process with standardized tools for evaluating proprietary names.

Second, we heard that all products made available to patients, whether they are prescription or over-the-counter drugs, should be held to the same standard of testing.

There is a need to try to simulate these kinds of situations that reflect real-life drug order situations to really evaluate in a realistic fashion the potential for problems in naming confusion.

Indeed, the study designs, to the degree they can, should replicate medication order situations where there are known error vulnerabilities.

And how medication orders, for example, are communicated can either be improved to reduce the potential for errors and how current medication order communication scenarios contribute to the propagation or continuation of those errors.

Particularly in the area of pediatrics, if one is looking at pediatric patients, it's important to not only look at confusions associated with the name, but also issues related to how well communication is managed in terms of the strength, the quantity, and the directions of use, as well as critical prescribing information, such as

patient age and weight.

There must be study methods that can be scientifically validated, reproduced, and that are objective and transparent to all.

One of the issues that was also raised at the public hearing, which we are not going to address today, is the issues of suffixes and prefixes associated with drug names which also have the potential and, indeed, to contribute to the problem of medication errors, nor will we be dealing today issues associated with over-the-counter family names and drug names that are marketed based on consumer recognition that lead also to consumer confusion.

So basically the major theme is that we feel that there is inconsistency in how name testing is currently conducted, that there is the need to produce valid and reproducible findings. You'll hear today that while all methods offer some value, we need to think about how to use these methods probably in a complementary fashion to come up with ways to prevent unneeded confusion once a product is marketed.

Following this open public meeting today, we will take both the results of the input we receive from the public as well as from our advisory committee, summarize these, as well as what we learned from June, and then look at the degree to which we can come up with a guidance to

industry that will provide them direction on how best to conduct pre-marketing testing and to communicate those results and data to the FDA.

Today following my presentation, we're going to hear from Jerry Phillips about the way we approach name testing at the Food and Drug Administration. We'll be hearing from a representative from PhRMA to talk about industry's approach, and then hear from five experts who are listed on the agenda talking about a variety of techniques that are currently being used to evaluate names.

We've asked each one of our expert panelists to provide an overview of each method, to discuss how that method should be validated, to determine how a study design can be used to evaluate how drug names can be studied to reduce medication errors, and the strengths and weaknesses of each of those methods.

Today we will consider the pros and cons of also taking a risk-based approach to testing proprietary names, to identifying the critical elements of each method to be included in good naming practices as part of a guidance document, to describe circumstances when field testing would be important and should be required to indicate whether one method should stand alone, and to describe circumstances when it would be appropriate to approve a proprietary drug name contingent on a risk

management program.

Thank you all very much and I will now turn the proceedings over to Dr. Gross.

DR. GROSS: Thank you, Dr. Seligman.

The next speaker is Robert E. Lee, Jr., Assistant General Patent Counsel at Eli Lilly and Company. He is going to talk on views on trademark evaluation. He's representing PhRMA.

MR. LEE: Thank you for this opportunity to share PhRMA views on pharmaceutical trademarks.

I would like to start with an echo from the June 26th, 2003 public meeting that PhRMA was honored to co-sponsor with FDA and ISMP. Among the points in my closing comments at that session was the observation that the role of trademarks in medication errors remains unknown. We do know that trademarks are part of most medication error reports, not necessarily as the cause, but as a convenient identifier for the products involved. PhRMA companies are interested as anybody in seeing medication errors eliminated. We believe that methods used by most PhRMA sponsors are an effective method for developing trademarks that help prevent medication errors. We are willing to work with the FDA and others on validated, improved methods, if it is possible that such can be developed.

Pharmaceutical trademarks are very visible and because they are so visible, they make an easy target for blame and criticism. The expression, "trademarks cause medication errors," has become an unchallenged part of regulatory language. Since PhRMA has not been able to find scientific support for the assumption, we think that this characterization is an overstatement and this is the time and place for it to be respectfully challenged.

Individuals inside and outside the FDA may unknowingly criticize trademarks when they use and overuse the expression "problem name pairs." For example, during the June 26th public meeting, Cozaar and Capoten were described as a problem name pair because they were involved in medication error. Cozaar and Capoten may have been involved in a medication error, but we do not agree that they are confusingly similar.

I have five points I'd like to cover this morning.

Point number one. Pharmaceutical trademarks support medication safety. The very essence of a trademark is to distinguish one manufacturer's goods from those of another. To do this effectively, trademarks must be distinctive and unique. It is this distinctiveness that serves to avoid confusion among current users and future users. This benefits both the manufacturer of the product

and the consumer of the product. Later on I will discuss in more detail the hard work that many manufacturers expend to develop pharmaceutical trademarks.

Distinctive and unique pharmaceutical trademarks support medication safety because there are no better product identifiers than trademarks. Nonproprietary names such as USANs and INNs use a stem system that is designed to group products together that have therapeutic class similarity. This creates a built-in similarity for generic names using the same stems.

Numbers would be a poor choice for product identifiers, and combinations of numbers and letters would probably be worse. Note that public internet addresses changed from the internet protocol addresses that used strings of numbers and letters to mainly alphabetical domain names that are easy to pronounce and remember.

As noted earlier, we are not able to find solid scientific data to show the role that trademarks play in medication errors, but it is easy to find public statements, news reports, and trade publications that echo the assumption that 12.5 percent of medication errors reported to FDA are a result of confusion between drug names. Yes, trademarks are involved in medication errors, but the involvement is most often in the convenient reporting of the errors, not the causes.

For example, the name pair Clinoril and Oruvail is among the several hundred problem name pairs listed in the USP Quality Review publication. Another pair among those listed is Cozaar and Zocor. We can all assume that well-meaning practitioners reported errors or near misses involving these trademarks, but we should not assume that these trademarks are so confusingly similar that they caused the problem.

FDA states that there are more than 700 problem name pairs, but only some of them contain two trademarks. Some contain a trademark and a generic name, and still others contain two generic names.

Rather than having the profession and public believe that trademarks cause medication errors, shouldn't we pause to perform a differential analysis to better understand the relative roles of the many factors involved in medication errors? PhRMA agrees that more work must be done to prevent or minimize medication errors. However, putting an inappropriate focus on trademarks, while ignoring other factors, gives a false sense of security that something significant is being done to reduce medication errors, while the underlying causes continue to put patients at risk.

Improvements at the prescription level are needed. One such initiative is legislation enacted in July

2003 in Florida that requires physicians to print prescriptions legibly. Another is similar legislation enacted by Washington State.

A number of promising improvements at the dispensing level were described by the late Dr. Tony Grasha at the University of Cincinnati. His research demonstrated that dispensing errors can be reduced by changes in the pharmacy work environment such as the use of prescription copyholders at eye level, limiting pharmacist workload, adequate lighting, improved equipment, et cetera.

These and other initiatives at the prescribing and dispensing areas hold promise to reduce medication errors.

Point number two. There is a highly effective method for developing pharmaceutical trademarks. The current method used by sponsors for developing new trademarks has been refined over the course of two centuries under the common law and trademark statutes. It is the most reliable method we know for determining whether two trademarks are likely to be confused by prescribers, dispensers, or consumers of the product.

During the early years, the central issue of likelihood of confusion was generally decided by comparing the various characteristics, similarities, and dissimilarities of the marks and the goods. But over time,

analysis of likelihood of confusion became more sophisticated and continues to evolve.

For example, in recent years most PhRMA companies seek input from health practitioners on the front lines so as to take into account various factors such as the frequency of prescribing, the consequences if products are mixed up, the dosage form, dosage strength, dosing regimen, delivery system, dispensing environment, the end user, et cetera.

Fact-based expert opinions made by trademark attorneys are also enhanced by continuous feedback from the judicial system. This judicial experience on issues of confusing similarity teaches us that the likelihood of confusion is a fact-driven expert determination. Similarity is a factor, but only one factor. Ultimately, trademark attorneys and judges apply many factors to all of the facts to reach a decision, and the decisions rest on the reliability and the relevance of the facts.

Through the research and writings of Dr. Bruce Lambert, we have some evidence that the industry is doing a reasonably good job of safely adding new trademarks to those already in use. Using various research tools to measure orthographic similarity, like trigram analysis, Dr. Lambert concluded that contrary to some impressions that the drug lexicon is getting too crowded, the evidence

presented suggests that most pairs of drug names are not similar to one another. This was in Dr. Lambert's paper, An Analysis of the Drug Lexicon.

Point number three, creative development and related activities. Creating distinctive and unique trademarks is a carefully constructed process that begins as long as four to six years before product launch and involves a great deal of sponsor resources.

There are some differences among sponsors, but the overall approach begins with creating long lists of candidates. These can come from internal resources or from outside vendors with extensive experience in trademark creation. It is not unusual for the initial list to contain several hundred candidates. These long lists are narrowed through an internal process where the emphasis is on eliminating candidates because they have potential safety risks or other problems. As the list is narrowed to a workable number of about 30 candidates that the sponsor believes are appropriate for the product profile, they are put through a more intensive screening process with increasing emphasis on similarity to other trademarks, generic names, medical terms, et cetera. Trademark candidates must survive the safety screens along with evaluations from legal, regulatory, linguistic, and commercial perspectives.

Trademark clearance is a detailed process that involves four stages, each of which weeds out candidates that have an unacceptable similarity to other trademarks based on an experienced analysis of the data. We not only compare candidates with trademarks that are on the market, but also those in the official trademark registration files in the U.S. and other countries around the world.

Stage one deals primarily with look-alike and sound-alike similarity and relies on search engines that are powered by sophisticated algorithms. For example, a typical approach is to sort trademarks by prefix, infix, and suffix using Boolean logic to combine letter strings into various configurations. This is an interactive process whereby the expert searcher changes the searching strategy depending on the results from the previous search run. This process continues until the searcher is convinced that the most relevant preexisting marks have been found in the database.

Another approach relies more on sophisticated phoneme analysis to measure phonetic similarity. Pat Penyak was going to be here from Thompson & Thompson to speak a little bit at the public session on what Thompson & Thompson does researching. Unfortunately, Pat was in an automobile accident, so she's not going to be here. I understand she's fine. I think there will be someone else

from T&T here today.

Comprehensive search reports are the raw data that is analyzed by trademark attorneys who perform an expert evaluation of similarity issues from both the visual and phonetic perspectives.

Stage two of the clearance process involves input from front-line practitioners who supply insights into how the trademarks will be used in a clinical setting. In addition to name similarity, the input from the clinical environment covers such elements as: frequency of prescribing, that is, popularity of the product; route of administration, dosage form, dosage strength, the usual regimen, clinical indications which hold important information about patient issues, storage, special preparation requirements, dispensing environment, generic name.

Stage three deals with forming the expert opinion. Once the searching and fact-gathering are complete, the sponsor team, comprising various disciplines such as legal, regulatory, clinical, and marketing, applies these various factors to all the facts available.

Pharmacists provide relevant input about the clinical and dispensing environment.

The legal searching provides insights into the look-alike and sound-alike similarity of other trademarks

with earlier priority rights.

Marketing and linguistic input identifies marks that are suitable for the relevant universe of prescribers, dispensers, and patients.

All of these inputs provide the resources for a fact-driven expert judgment about the suitability of the trademark for use on the product under consideration. It is only after all of this work is completed and all the results reviewed that a decision is made on which trademark, among the few survivors, will be adopted and moved to the next stage.

Stage number four, the final stage in the process, involves the filing of an application for registration in the U.S. Patent and Trademark Office. Even with all the searching and fact-gathering that formed the basis for the selection decision, there are more reviews and hurdles ahead. Typically all pharmaceutical trademarks are filed in class 5 at the Patent and Trademark Office. This class contains more than 150,000 applications or registrations in the U.S. alone, more than a million worldwide.

PTO examiners who are experienced in reviewing pharmaceutical trademarks conduct an independent search of the candidate trademark for confusing similarity. These examiners, working in class 5, apply a higher standard for

pharmaceutical trademarks due to public health concerns.

If the examiner finds the trademark acceptable under the PTO review standards, the trademark is published in the Official Gazette, a weekly publication that contains all trademarks recently filed. Competitors and others routinely review the Official Gazette to see if any of the trademarks published might be unacceptably similar to their own marks.

If a published trademark is determined to be unacceptably similar to the owner of the trademark with a priority right at the PTO, the owner can file a notice of opposition which stops the PTO approval process until the opposition is resolved by adjudication or settlement.

In a situation where an issue of confusing similarity arises between two trademark applications, it is necessary to determine who has the right to register the mark. In the U.S. and all other countries, trademark laws provide that the first to file an application has priority over the later-filed trademark application.

The national trademark systems are tied together by treaty so that priority is assigned to the first filed application in any one of the treaty countries. This is an important matter and has legal implications if overridden by a priority scheme not endorsed by Congress.

Point number five, promise and pitfalls of

computer technology. We learned that FDA is working with the Project Performance Corporation to develop a web-based drug comparison system called POCA, an acronym for Phonetic and Orthographic Computer Analysis. New and improved software tools and databases can support the process of trademark selection. PhRMA looks forward to being part of the development of the new software so that it can be integrated into work being done by commercial vendors with similar interests.

We do see some serious pitfalls with the POCA project. The first is the fear that FDA would not openly share the system with sponsors. We think it is important for sponsors to have the option of integrating any new FDA-sponsored software into existing trademark evaluation processes. The second is the fear that FDA would use output from POCA to second guess the decisions about trademark acceptability made by sponsors who follow the processes that I described earlier.

Recommendations. In closing, I would like to make four recommendations.

One, FDA should recognize the intrinsic value of trademarks that make it possible for billions of prescriptions to move through the dispensing and administration process error-free. In addressing the small percentage of prescriptions that result in medication

error, FDA and others should focus resources on the major unaddressed causes of these errors.

This is number two. For all the reasons I've given today, FDA should recognize the value of the current methods employed by sponsors to develop clear and adopt new trademarks for pharmaceutical products as an effective working model of good naming practices. The current process includes review and judgment by front-line practitioners, the sponsor trademark attorney, the PTO examiner, and competitors before a trademark is adopted. Careful consideration should be given to the extent of further trademark review by FDA so as to avoid moving beyond the point of diminishing returns.

Number three, FDA has an interest in making sure that pharmaceutical product names are chosen with care and should exercise its regulatory leverage in seeing to it that sponsors select trademarks carefully. FDA should establish guidelines, based on the sponsor process described earlier and insure that the guidelines are followed.

FDA should encourage the development of improved computer software tools, more comprehensive databases, and additional research so long as FDA recognizes that the process for determining the suitability of a new trademark is largely a fact-based expert judgment

that should be made by those who have the professional expertise.

Thank you for your kind attention, and I'll be here all day for any questions.

DR. GROSS: Thank you very much, Mr. Lee.

Next we will hear from Jerry Phillips who is Associate Director of Medication Error Prevention at the Office of Drug Safety. He will present the FDA's approach to proprietary name evaluation.

MR. PHILLIPS: Thank you. I'm going to talk a little bit about a couple of things. I'm going to give some definitions. I'm going to tell you a little bit about our perspective as far as the seriousness of the issue and then our process for evaluation at FDA.

First, let's start off with the definition of a medication error. This definition comes from the National Coordinating Council for Medication Error Reporting and Prevention and it has also been proposed in the SADR rule by FDA. Basically the key word here is that it's a preventable event that may cause or lead to inappropriate medication use or patient harm while the medication is in the control of a health care professional, a patient, or a consumer.

FDA focuses on medication errors that relate to the safe use of a drug product. In its perspective, that

includes the naming, the labeling, and/or packaging of a drug product that might contribute to an error.

A proprietary name by definition is a name that's owned by a company or an individual and is used for describing its brand of a particular product. It's also known as a brand name or a trademark.

We just heard some of the statistics on the 700 name pairs. I acknowledge that both proprietary and generic names are part of that list. Some of those are actual errors and some of them are potential errors that are on this USP list of 700 drug names.

To date about 25,000 medication error reports have been received by FDA. When we look at the database, we do a root cause analysis of those events and determine the causes of those. From the aggregate data, approximately 12.5 percent of the errors are related to the names. This is from the reporter's perspective of the cause of the event.

FDA, myself and others on the staff, publish mortality data that was collected from 1993 to 1998 and was published in the American Journal of Health System Pharmacists on October 1, 2001. Of this data, we had 469 fatalities due to medication errors. A breakdown of this is 16 percent of the deaths were due to receiving the wrong drug product. Now, receiving the wrong drug product

doesn't mean it's necessarily related to the wrong name. A physician could write for the wrong drug and that product could be administered. But if we look at proprietary name confusion and generic name confusion, 5 percent of the deaths were caused by proprietary names and 4 percent by generic names.

There are many, many causes of medication errors such as lack of communication, use of abbreviations, handwriting, lack of knowledge. There are many, many reasons.

Some of the other reasons include similar labels and labeling. In this particular picture, what you see is a blue background. You see red lettering. You see a standardized format on these particular bottles, and this can lead to selection errors.

In this particular case, these are ophthalmic drug products manufactured by one particular company, and you can see the similarity across the different products that increases the chance for selection errors.

This is an example of an over-the-counter drug product. This is that OTC family trade name issue that we're not going to talk about today. But basically it's a similar labeling and packaging. These two drug products have different active ingredients. One is oxymetazalone. The other one is phenylephrine. They both have different

durations of action, and it has led to confusion.

Names that don't seem to be similar, Avandia and Coumadin, when written sometimes do look very, very similar and have resulted in errors. This is an example of a prescription written for Avandia 4 milligrams every day and Coumadin 4 milligrams every day. The similarity, having both identical strengths, both being written for every morning increases the risk of a medication error when these names are written together and have resulted in errors.

So what is FDA looking for when we look at trade names? There are basically two things. We look for sound-alike/look-alike properties of that name and we also look for promotional and misleading claims associated with that proprietary name.

For sound-alike/look-alike properties we're looking at currently marketed and unapproved drug products that we have in the pipeline. We're also looking to other medicinal products and to commonly used medical abbreviations, medical procedures, and lab tests.

So what's the information that we need in order to do our risk assessment? Of course, we need to know the proprietary or trademark and its established name. We also need to know how it's going to be dosed, its strength, its dosing schedule, its use and its indication, its labels and

labeling. If there's a device involved, we ask for the working device model, and we also look at the formulation and the packaging proposed, along with the trademark.

This is a busy schematic flow of the process at FDA. There's a request for a proprietary name consult that comes from the product sponsor, and that is at any time from phase II of an IND to the filing of the NDA, the sponsor requests the name through that IND or NDA, and it is then filed in the reviewing division. A project manager will consult the Office of Drug Safety or the Division of Medication Errors and technical support in that office.

The review, which I'll go into a little bit more detail, is a multi-faceted review that starts off with an expert panel. We use computer analysis, POCA, which was mentioned earlier, and prescription drug studies. Then a risk assessment by a safety evaluator on DMETS's staff is done that takes into account all this data. The review goes to a team leader, a deputy director, and the associate office director. Recommendation is then given back to the reviewing division who reviews our consult. They either agree or disagree with it and then provide that information back to the sponsor.

As I just mentioned, the analysis consists of an expert panel, a computer analysis which looks at the orthographic/phonetic similarities of a name. We search

other external computer databases. We perform prescription drug studies. These are simulated prescription studies that try to simulate the real world as far as prescribing practices, which include a verbal order, an outpatient written prescription, and an inpatient written prescription. And then we provide an overall risk/benefit assessment based upon the information that we've collected.

The expert panel consists of approximately 12 of the DMETS safety evaluators. This includes a physician, pharmacists, nurses, and one DDMAC representative. That's for advertising that renders an opinion for misleading or promotional claims.

There is a facilitator in this expert panel that is randomly selected and rotated.

Each expert panel member reviews reference texts, computers, and provides a relative risk rating for each name prior to the meeting.

Then there is a group discussion at the expert panel and there's a consensus that's built on each particular name.

From this, we design prescription drug studies. From the expert panel, there may be several names that have been identified by those experts of marketed drug products that might be confusingly similar. And from that, we design these studies where we will write an outpatient

prescription with the proposed name and an inpatient prescription written and also a verbal order.

The prescription study designs are developed specifically for failure mode. In other words, we stress the tester by randomly selecting different types of handwritings, using actual practice standards. Instead of putting an indication on a prescription, we would leave that indication off because putting the indication on necessarily doesn't reflect normal current practice and it would also lead the analysis in a different direction so that you wouldn't get an error necessarily.

We have various staff members that are asked to write sample prescriptions for each name. There is a marketed drug or control prescription that's also included in the prescriptions so that the tester knows that they're evaluating unapproved drug products, but also we'll put in some marketed drugs. Sometimes we'll include marketed drug products that are known error pairs to validate the prescription studies.

The prescription is scanned and then they're e-mailed to a subset of FDA health care workers. Their interpretations are e-mailed back to us in writing.

There are about 130 FDA physicians, nurses, and pharmacists across the centers that respond by this e-mail system with their interpretations and comments. To

eliminate any one reviewer from reviewing a name more than once, we divide the entire group into thirds where the n is approximately 43 to review each verbal order, written, and outpatient prescription order. The response rate is usually around 70 percent.

This is an example of a product that we had on a scientific round. This was not a proposed name by a drug company. It was called Novicar. The top prescription is an example of the prescriptions that we normally scan for our participants. In this case, we had written out the patient's name and the date, Novicar 40 milligrams, 1 PO every day, #30, and Dr. Opdra at that time.

The bottom is example of an inpatient order that we wrote for this study that gives the diet of the patient, blood work, a DC order, and the Novicar is put in there also. The lined orders on an inpatient order present different types of errors because of the lined orders, and that's why we duplicate both.

Just to back up, on this particular study we actually discovered that there were lots of errors with Novicar with -- oh, shoot. I just forgot. I'll come back to it.

VOICE: Narcan.

MR. PHILLIPS: What was it?

VOICE: Narcan.

MR. PHILLIPS: Narcan.

On verbal orders, randomly selected DMET staff are asked to record a verbal prescription via telephone recorder. An example. This is Dr. Dee Mets and I'm calling in a prescription for Jane Doe for Novicar 40 milligrams. I want to give 30 with two refills. And that's recorded and then sent to the group of physicians and nurses and pharmacists on the prescription drug studies. Then after they hear that, they e-mail us back their interpretations.

We also use a phonetic and orthographic computer analysis. This is a recent software that we have contracted. We abbreviate it as POCA. It's a set of phonetic and orthographic algorithms that are used for an automated and computerized method for evaluating trade names for their similar sound-alike and their look-alike properties. The prototype has been completed and is in operation currently and is being used routinely in DMETS's reviews. We are also working on validating this prototype and hope to have that completed soon.

POCA provides a percentage ranking of orthographic and phonetic similarity between the proposed name and the database of existing trade names that it compares itself to. It also considers the similar strengths and dosage forms when looking at a name.

Now, the safety evaluator also does a risk analysis and they examine the data from the expert panel that was originally done, the prescription studies, any computerized searches, POCA to establish any risks for confusion. They also evaluate the potential safety risk associated with two identified drug products being confused with each other due to that similarity and examine their post-marketing data -- that's preventable adverse drug event data -- their clinical and regulatory experience and any literature reports. It's important to take the lessons that we've learned from post-marketing into this evaluation also.

Some contributing factors for name confusion include similar indications, having the two drug products prescribed in the same patient population, having identical formulations, overlapping strengths or directions, being stored in the same area.

We also look at what's the potential for harm when we look at the two trademarks. What are the consequences if a patient misses the pharmacological action of the intended drug? We ask these questions routinely. And then we ask, what are the pharmacological actions and toxicities of the unintended drug product?

There is a final review done. There are actually basically two reviews that are done on trade names

at FDA: first, the initial one that I just described which was a multi-faceted review, and a final review that's done approximately 90 days before the action on the application. We don't repeat the extensive evaluation that I just mentioned. We're only looking for any confusion with names that have been approved since the initial review was done and to the time in which the application is going to be approved for FDA approved names during that interval.

I thank you very much.

DR. GROSS: Thank you, Mr. Phillips.

The next speaker is Dr. Bonnie Dorr, Assistant Professor, Department of Computer Sciences at the University of Maryland. She will talk about automatic string matching for reduction of drug name confusion.

DR. DORR: And make that Associate Professor.

DR. GROSS: Congratulations.

(Laughter.)

DR. DORR: It's seven years ago now. Thanks.

So I'm going to talk about automatic string matching, some of the things that you've heard already that are part of the technology behind POCA, and I'll also talk about other analyses that are done that, combined with some of that technology, could potentially get improved results.

So these are the questions, just to remind you, that we were asked to address. I will be giving an

overview, some of which you've probably seen before -- but it never hurts to review -- of phonological string matching for ranking. Also, I will be looking at orthographic string ranking.

And validation of a study method. What we use is precision and recall against a gold standard to determine the effectiveness of the different matching approaches.

I'll talk about an optimal design of a study, and interface for assessing appropriateness of the newly proposed drug name.

And then finally, strengths and weaknesses. Each algorithm can miss some correct answers and also get too many that may not be appropriate. So we'll learn more about that.

So this is the overview. String matching is used to rank similarity between drug names through two different techniques. Some of these were mentioned. Orthographic compares strings in terms of spelling without reference to sound. Phonological compares strings on the basis of a phonetic representation or how they sound. Within those, each of them has two different types of matching that are done. One is by virtue of distance. How far apart are the two strings? And the other is by similarity. How close are the two strings? If two drug

names are confusable, of course, we want the distance to be small and the similarity to be big. So that's the basic idea.

I'll give some examples briefly of different orthographic and phonological approaches, both with distance and similarity.

Under the heading of orthographic, we have a couple of distance metrics that are actually related, the Levenshtein distance and the string-edit distance. There's a function between those, so they come out to be about the same when you do an analysis.

I'll talk about LCSR which is the Longest Common Subsequence Ratio, and Dice. The LCSR and Dice are similarity metrics, all under the heading of orthographic.

Under the heading of phonological, I'll talk about a distance metric that is based on sounds called Soundex that's been around for a long time versus a similarity metric under the heading of phonological called ALINE. You may see some typos floating around. Sometimes it's spelled A-L-I-G-N, but this is actually the name that was used for the system.

When we want to compare distance and similarity, we want to sort of look at, okay, what do you mean how far apart or how close? Can I look at those two and say whether there's a relation between them? Usually

what you do is you say the distance between two strings, two drug names, is comparable in some way to 1 minus their similarity. It's the number between 0 and 1, so if you subtract it from 1, you get a number that allows you to compare these.

Orthographic distance. Essentially with the Levenshtein and string-edit distances, you're counting up the number of steps it takes to transform one string into the other. Some examples are given here where, as you can see, the bold-faced pieces here indicate the places where the two strings are different, and the remainder is the same. So you're actually counting the number of places that you're different. That's the Levenshtein or string-edit distance.

Also, if you look at Zantac and Xanax, you can see that the X's are counted as different. Even though certainly the initial X sound sounds the same as the Z at the beginning here, they're taken to be different. So the number is 3. Then typically what we do to get sort of a global distance is we divide by the length of the longest string. So we actually know that this distance is really .33 because you have to factor in the length of the string as well; whereas, for the latter one, you're talking about a distance of .5. This is actually a counterintuitive result. If you use Levenshtein or string-edit, Zantac and

Xanax are more distant than Zantac and Contac, and that's not a result that you want. So we'll talk about that.

LCSR. In this approach, you double the length of the longest common subsequence and divide by the total number of characters in the string. What does that mean in terms of these same examples? You're looking at the similarity in this case, because before we were looking at distance, so we were highlighting the Z and the A. Now we're actually going to highlight the rest of the string. We're going to look at where they're the same. We're going to do a doubling operation here. That's 2 times 4. We're going to divide out. We get .67 here, whereas with Zantac and Xanax, highlighting the characters again that are the same, you get .55. Now, in this case this are reversed. You're talking about similarity. So we're actually in this case saying that Zantac and Contac are more similar than Zantac and Xanax, which also is not a result that you want to get.

Dice doubles the number of shared bigrams. What are bigrams? That's just two characters that occur together, and you divide by the total number of bigrams in each string. Some examples are shown here. If you take Zantac and you sort of pull out all its bigrams, and then Contac and pull out all its bigrams, and then you do this doubling operation again, you divide by the total number of

bigrams in each string, you get .6. Whereas, if you do the same thing with Zantac and Xanax, you're going to get .22. Again, these are similarity metrics which means you really kind of want Zantac and Xanax to be close, and they aren't close. They're .22 compared to Zantac and Contac which are actually .6. So, again, we're getting a result that we don't particularly want. But these are common techniques that have been used in the literature.

Another technique, now moving to the phonological approaches, moving away from look-alike and getting into sound-alike. Here what you do is you transform all but the first consonant to numeric codes. You delete 0's and truncate resulting string to four characters. This is a character conversion that's referred to here. You're actually sort of mapping the vowels to nothing. The 0 means they just drop out. These consonants here kind of sound alike, so they get a 1 and so on. So each of these sets of consonants is going to get a particular number.

To give you some concrete examples to work with, this allows you to say "king" and this sort of version of "khyngge," sort of an archaic version. They sound alike and they each get the same code: k52, k52. So those, indeed, look the same.

Unfortunately, if you really apply this

thoroughly, you get "knight" and "night" aren't the same because one of them is k523 and the other is n23.

And even worse, things like "pulpit" and "phlebotomy" come out to be the same when they are radically different, and so you get some pretty bad results there.

So the same thing with Zantac and Xanax. You're missing out on that commonality between the initial Z or X sound.

Also, an alternative approach to sound-alike that has been used that's been reported in the literature is to compare, instead of using phonological distance of this type, the syllable count, the initial and final sounds, and the stress locations. But this has been shown to miss out on some confusable pairs like Sefotan and Seftin because that has a different number of syllables, and Gelpad and hypergel, where you sort of swap things around, and "gel" is at the beginning of one and at the end of the other.

So really, what you need is something to provide that -- the pronunciation for sound-alike -- you need to be able to capture what's going on there for those types of similarities. So ALINE is something developed by Greg Kondrak in the year 2000 to use phonological features for comparing words by their sounds. Some characters are

missing here but it doesn't matter much. Those two lines right there are telling you that an ending X sound sounds like KS as in Xanax, but and initial X sound sounds like Z. So if you take those and break them down into the features of what those phonological symbols mean, really you can talk about the pronunciation, the position of the tongue in the mouth and where it stands with respect to the teeth and the back of the mouth, and that's what those features mean in here, without going into detail.

The point is that you're going to use, instead of a part of a string as in Soundex, the entire string. Instead of dropping vowels as in Soundex, you're actually going to keep them and they are going to be more significant in drug names. And you're going to use decomposable features in determining the sorts of confusions that people get.

This was developed originally for identifying cognates and vocabularies of related languages such as "colour" versus "couleur" in French. But the feature weights can be tuned for a specifically application, which is what we've done with this system.

In this approach, phonological similarity of two words is reduced to an optimal match between their features. So what we do is we take something like Zantac and Xanax and we align the characters by virtue of going

through the decomposed features of this form.

Just to show you another example. This is Osmitrol and Esmolol. This is a schwa. It's missing. It isn't missing in mine, but they don't always port over to other people's machines.

So the approach that's being used here is to sum up the weight of the match on each sound. In fact, you can align the characters of the strings by looking at their underlying phonological sound. The E in the Esmolol is actually a sound. You take an alignment and you balance out across the features of each of those. If you've got a good match, you get a higher score. So the M and the M get a very high score. In fact, that's a maximal score, whereas this vowel sound in here is close. It's certainly higher than a 5, but it's not up to a 10, and so on. And then you add up and you get a 58 here, and then you normalize it by the total maximum score which would be 80 in this case. You could get a potential score of 80 if they were identical strings to get a number like .73.

So this approach identifies identical pronunciation of different letters like the M that we saw. It also identifies non-identical but similar sounds such as this one at the head of the two words.

Of course, I have to show you a picture of a head with a tongue and teeth, just to make sure that you

know that I'm a computational linguist. But the idea is that there are positions within the mouth that -- sound is produced through the vocal tract and also involves the position of the lips, the tongue, the teeth, the hard palate, the soft palate. That's all called place of articulation. Everything bundles up under place of articulation. But also the manner in which air passes through the oral cavity which we call manner of articulation. So there are a lot of other features too, but the top two that we really like to focus on are place and manner.

These are some examples of places of articulation. So here is where the two lips are together. That's called bilabial. Here's where the tongue is right behind the teeth like a D or a T. That's alveolar, and so on. Here's a K sound where the back of the tongue is raised. This is called place of articulation.

And we can assign particular values. Each individual value within that feature is given a particular weight. So bilabial is really important for drug name matching, for example, and the other ones may be less important.

I said place of articulation and manner of articulation. There are also some others that I won't go into. These two are the heaviest weighted values. We

really focus on those and give them the highest score if we get a match there.

Just to give you some examples. So these are showing the Zantac/Contac comparison that I gave you earlier with Edit, Dice, and LCSR. I already had given you those scores and I showed how they were computed. In the case of ALINE, we actually have Zantac and Xanax as the highest scoring pair out of the three different pairs, the three different combinations that you can get, which is much closer to what we would like to see. We'd like to see that we're looking at the initial sound as something that humans consider to be phonologically equivalent even if the characters are different. So that one actually gets a higher score, whereas Zantac and Xanax in the others do not get the highest score, come in sort of second place.

Question number two was how do we validate this approach, and the answer for this is to use something called precision which is counting up the number of matches your algorithm found. We could try this with Edit, Dice, ALINE and so on. Take each one of those algorithms, count up how many matches that it got, and take that over the number of correct matches that you could possibly get, and that's precision.

Recall is the number of correct matches in your problem space versus how many does your algorithm determine

to be a match. So that's the notion of recall.

We use the USP Quality Review as our gold standard. This is necessary in order to determine precision and recall. There were 582 unique drug names, 399 true confusion pairs, and if you multiply these out, combinatorically you could get 169,000 possible pairs. You can then rank all of those pairs according to -- in this case I'm not using ALINE. I just put Dice up here. You could rank them according to whether they match with that particular algorithm.

So Atgam and ratgam was the one that came out the highest. Using Dice, it came out with a score of .889. It has a plus sign in front of it, which means it did occur in the USP Quality Review as a confusable name pair. It also was the top ranking one.

Our next ranking one also has a plus sign, which means it did occur in the USP Quality Review as a confusable pair.

The next one down did not occur in the USP Quality Review but maybe it should. It looks like it's a typo. But in any case.

Quinidine and quinine. I'm not an expert on pronunciation of these particular drugs, but that was the next one down, and it did occur, and so on.

So you can figure out on the basis of these,

and how often you're getting the correct answer out of your gold standard, what your precision and recall values are. If you map that out, the way to do it is to compare precision at different values of recall. So the precision is along this axis. How precise are you being with your answers? How many correct answers are you getting? Over how many correct answers out of the problem space are you getting. If you take those two together, you get a graph that looks like this. ALINE is the top score over here with the sound-alike version.

If you turn ALINE into the look-alike version -- there is a version that you can just take out all the pronunciation -- it still gets a pretty high score. In fact, it even gets higher than the sound-alike version in one place. But they look pretty much the same for several values of recall, whereas LCSR is lower-performing. Edit is the blue line here, and Dice is down here.

At least we have a feel for the idea that somewhere in this manner and place, the places of articulation in the mouth, the way air passes through the mouth, is doing something to get us closer to the USP Quality Review, with the caveat that there are a lot of other errors recorded in the USP Quality Review, of course. In fact, we had to do some studies that are not reported here on cases where it wasn't such a large list of many

names that people had speculation and other things factoring into it. So we worked with another list as well and got similar results, but I haven't brought that in here.

We really do need to make sure of the transcription into the sound form isn't what's getting the full power of our matching. That is, if we gave Dice and LCSR that same ability to look at sound, would they perform as well as ALINE. It turns out they don't. The sound and the non-sound versions of Dice and the sound and the non-sound versions of LCSR perform lower than ALINE with its phonetic transcription. There's something going on with the weighting and the tuning of the parameters based on articulation points that gets us the higher value.

So what would an optimal design of a study be? I actually agree with Dr. Lee that a system should be openly shared, that an optimal study would involve the development and use of a web-based interface that allows applicants to enter newly proposed names. That same software should be used by FDA to ensure consistency of scoring so that everybody is looking at the same scoring mechanism. And that design would ensure that updated versions of software would be continuously available to potential applicants.

So the interface would display a set of scores

produced by each approach individually, as well as combined scores based on the union of all the approaches. That's something I want to get into. Even though ALINE is the highest-scoring one, there are reasons to look at the combinations of the different approaches to figure out the best answer.

The applicant could compare the score to a pre-determined threshold to assess appropriateness, or that threshold could be set community-wide.

In advance, running experiments with different algorithms and their combinations against the gold standard would help to determine the appropriateness for the threshold and also allow for fine-tuning, calculating the weights for the drug name matching.

Just continuing along that last point there, right now the parameters have default settings for cognate matching, but they may not be appropriate for drug name matching. Something that we might want to do as a part of this is to calculate the weights for drug name matching and then use hill climbing to search against a gold standard to get the values that we're giving for the articulation points closer to what we need for drug name matching.

For our initial experiments, we did tune the parameters for the drug name task, looking at things like maximum score, which has to be a high threshold for cognate

matching, but should be lower for drug name matching because we ended up with things where it was too risky to consider certain pairs to be the same. Like the "puh" and the "kuh" sound should not be considered the same for drug name matching, whereas in cognate matching, they should be. Also there was something called an insertion and deletion penalty, which should be low for the cognate task but higher for drug name matching. Because confusable names are frequently the same length, a vowel penalty which for cognates, the vowel penalty is low. Vowels are less important than consonants, but that's not true of the drug name matching. Again, we're taking this from a field and moving it into a whole different application, so this type of tuning is necessary. Phonological feature values for drug name matching, place distinctions should be ranked as high as manner distinctions.

Last question. Strengths and weaknesses. Just sort of repeating something Dr. Seligman said, all methods offer value and should be used complementarily.

So here are some ALINE matches. ALINE gets these sort of pairs, but others don't because ALINE doesn't care whether there are shared bigrams or subsequences. It really is looking at the phonetic features associated with these. Again, these are pairs that I took out of the USP Quality Review.

On the other hand, Dice matches with these particular pairs, but others don't because Dice is able to match pairs of words that are similar with bigrams. If it can find that the S and the I is here and the S and the I is here, it's looking at that sort of thing. So ALINE would potentially have trouble with that. And it can do that even though the remaining parts are not the same. So gel and gel show up here, but the remaining parts are not the same, but Dice gets those.

LCSR gets these, but others don't because the number of shared bigrams is small for these types of pairs, Edecrin and Eulexin. I'm sorry for the pronunciation that I'm giving. Except for the "in" right here, there are no shared bigrams in this particular pair, but LCSR is able to find that as a potential confusable drug name pair.

Just to elaborate on each of those really from the previous slide telling you what's going on, ALINE, using interpolated precision, gets the highest score. It's easily tuned to the task and matches similar sounds even if there's a difference in initial characters like Ultram and Voltaren, but it misses words with high bigram count, as I mentioned.

And potentially the weight-tuning process may induce overfitting to the data, so if we get it trained up so that it gets this pair here, it may also get a false

pair, the Brevital and ReVia pair which is not one of the confusable ones.

Dice, on the other hand, matches parts of the words to detect confusable names that would otherwise be dissimilar, like Gelpad and hypergel, but misses similar sounding names like the ones that ALINE can get, the Ultram and Voltaren pair with no shared bigrams.

LCSR matches words where the number of bigrams is small like this pair I showed you on the last slide, but misses similar sounding names like Lortab and Luride that have a low subsequence overlap.

So the previous slide showed the weaknesses and strengths, but we think that taking a combined approach -- and in fact, we have some initial experiments from the last week or two that are not shown here, that the best approach is to use a combination of all of these to get closest to the gold standard. So we want to continue experimentation with different algorithms and their combinations against the gold standard.

Fine-tuning based on comparisons with that gold standard. So, of course, we still need to look at reweighting phonological features specifically for the drug naming task.

We believe that taking the phonological approach that has been designed in ALINE by itself and also

in combination with other algorithms provides a strong foundation for search modules in automating the minimization of medication errors.

And again, just reiterating that a combined approach that benefits from the strengths of all the algorithms, increased recall, without severe degradation in precision, that is, the false positives, is the way to go in my opinion.

DR. GROSS: Well, thank you, Dr. Dorr, for clarifying that confusing field for people who aren't in it.

(Laughter.)

DR. GROSS: We have time for some questions. Brian.

DR. STROM: I have three questions for Jerry. We heard from Mr. Lee that there wasn't a problem. We're hearing from you that there is. Let me ask each of the three separately. How often do you get a name from industry that FDA ends up rejecting?

MR. PHILLIPS: We reject about one-third of the trade names, and we review about 300 names a year.

DR. STROM: Second. How do you know which one was correct? In other words, were they correct in originally thinking it was safe, or was FDA's approach correct in rejecting it?

MR. PHILLIPS: That's difficult. I have case examples where we suspected problems of a drug name prior to approval, and for reasons, it got approved, and sure enough, we had post-marketing data that confirmed our opinions. I also have evidence that things that we had concerns about got into the marketplace and we never saw that come forth. So it's difficult to know who's right and who's wrong at times.

DR. STROM: A third question which is related. Dr. Dorr just gave us an elegant presentation versus a gold standard, the gold standard being the USP list of names. Why is that a gold standard, and what does that list represent? Clearly the idea of testing these methods against a gold standard make enormous sense. What I'm questioning is how gold is the gold standard?

MR. PHILLIPS: Well, the gold standard is from the reports that the USP has received of medication errors associated with both generic and trademark confusion. So that list is a representation of all the reports that have come in. Some of those reports are potential errors and some of them are actual. So the gold standard probably should be applied to those errors that occurred with trademark confusion pairs that actually occurred in an error and not a potential error. That's the reason why we chose that as the gold standard because it's actually based

upon actual clinical experience of people being injured or being involved in an error with those names.

DR. GROSS: Michael Cohen.

DR. COHEN: Thank you. I have a few questions too for the different speakers. I'll ask them as quickly as possible.

First for Mr. Lee, as you know, ISMP actually contributes to the FDA Medwatch database as well. The USP and ISMP together we actually have received many, many error reports with trademarks. I agree with you. They're always multi-factorial. There are many contributing factors besides the drug name. But would PhRMA acknowledge that at least one of the contributing factors clearly might be a trademark? Otherwise, how could you explain a change in a trademark totally eliminating the problem? For example, Losec and Lasix. It's gone. We never had another problem with that. Levoxine, gone when the name was changed to Levoxyl. So from that standpoint, I need that clarification to make sure that we're on the same page here -- the committee, that is, and PhRMA.

MR. LEE: Yes, I think there are certainly examples of name pairs on the marketplace that are more similar than others, but I would think the modern day practice, let's say, by PhRMA companies takes into account the clinical settings. I think with that screening with

the clinical settings, we should see less occurrence of the kind of name pairs like Lasix and Losec.

DR. COHEN: A second thing. This is for Jerry I guess. I wanted to know if he would acknowledge -- I agree with Bob and you -- I don't agree with you that the percentage of errors related to trademarks in the FDA Medwatch database is actually a true reflection of what's happening out there, and I think that should be pointed out because really what it is I think the reporters characteristically see FDA as a repository or an organization that can effect change with product-related issues. So the types of reports that you would get I think more than practice-related issues would be product-related issues and the kinds of things that you would get reported would be things that practitioners who report to the program think can be addressed by FDA. So I just wanted to point that out. We do see that figure quite frequently and it could be misleading unless you use it correctly, which is what you did, you said reported to FDA. You didn't say that's the actual percentage out there.

MR. PHILLIPS: I acknowledge that. That's the data based upon what we've received, and we have a system that collects data on drug products and more serious adverse events. So it is skewed in one direction.

I would mention that Medmarx has released its

annual report this year. I think there was some 8 percent of their reports of 192,000 reports that had something to do with name confusion. Some 4,000 patients were involved in errors. So I think there is some evidence outside FDA's reporting system that it still is a problem.

DR. COHEN: I'm not trying to minimize it. I'm just saying that it may not be 12.5 percent.

The other thing, for Dr. Dorr, I had two quick questions. Do you think systems like yours could be used as a sole method for testing?

DR. DORR: I don't know if you mean the technique, the methodology.

DR. COHEN: Yes.

DR. DORR: Right. So what we're experimenting with right now -- we actually have a pretty good result -- is bringing in a combined version of Dice, ALINE, LCSR, and so on. By the way, this is only for look-alike and sound-alike. So we have an orthographic version of it and we have a phonetic version of it. So we don't pretend to try to -- I guess that was 16 percent or 12 percent somebody said of the overall problem. So I agree with your comments about the USP Quality Review as taking in too many things that have nothing to do with that type of matching.

But I believe that taken alone, the phonetic approach, if you had to choose one, is the best one. We've

got some definitive, repeatable results on that. But you can get better than any of the approaches alone, including ALINE, if you take a combination of the different algorithms.

DR. COHEN: Then finally for you, what databases do you actually use?

DR. DORR: The only one was that USP Quality Review.

DR. COHEN: I see.

DR. DORR: Yes. Although more recently we have looked at something that was a proprietary database. I'm working with PPC, and so they had given us a smaller version of just names that are not in this sort of broader category of any medication error. And we were getting similar results on that one, but I couldn't put any of that on the slides.

DR. COHEN: Thank you.

DR. GROSS: Robyn Shapiro has a question.

MS. SHAPIRO: Yes. I still am somewhat confused about the underlying assumption, being a newcomer to this whole topic. To me the data about the causation is very weak. For example, Dr. Phillips, in your comments, the 12.5 percent by reporter, is the reporter always the individual who we think is responsible for that error? And if not, then how good is that data in and of itself?

And the confusion about the underlying assumption is important not only for us to kind of think about why we're here, but also where we're going. In other words, if a risk management approach really had to do with how we see these prescriptions written out, then the transcription would be the subject of our focus as opposed to the actual name.

So I'd like to know from the FDA how confident you feel about the causation of these med errors being attributable to the name itself.

MR. PHILLIPS: I feel pretty confident about the data that I have and the causation, that there is a contributing factor with similarity of trademarks, that they can definitely be associated with the event. There may be other contributing factors, but there is a definite association between similarities of names that contribute to errors.

MS. SHAPIRO: Based on data? You feel confident because you have data about that?

MR. PHILLIPS: That's correct.

MS. SHAPIRO: Could we see it?

MR. PHILLIPS: Within our Adverse Event Reporting System and the data that I cited, the analysis that was done over the 6-year period?

MS. SHAPIRO: Yes. Again, I'm interested in

pulling it apart so that we know, if we can, that these errors we feel confident are on account of the name as opposed to all these other factors that go into med errors. That would help me to think about a risk management approach.

MR. PHILLIPS: Usually when a reporter reports on a medication error, they're going to give a narrative of the event itself and usually will provide some causes of that event. That doesn't necessarily mean that reporter is correct. The reporter may not actually be involved in the error, as you cited. They may be reporting the event. A risk manager may be reporting the analysis that was done at a facility, and according to that facility, these were the contributing factors associated with that medication error. There are always more than one factor involved in an error. So just to say that it was just trade name was probably not true for the whole event. But if you do look at the narratives in the cases and look at these -- and you can run those similarities through an analysis yourself, and we do that -- you will see the similarities and the contributing factors.

DR. GROSS: We have three more questions. I'm taking more time for the discussion because it's beginning to get at the crux of the problem. Ruth Day.

DR. DAY: I have a couple of questions for Dr.

Dorr. First of all, you're comparing across these different computational linguistic methods. They all have their strengths and weaknesses, and taken together, they do a lot. It's great to see.

I'm concerned, however, they all depend on an initial phonetic transcription. So one part of that is who does the transcription. I have seen within companies, as they go forward with a given name, there are alternative pronunciations even within the company. We heard from you this morning quinine. Others say quinine. You could also say quinine and so on. So you might say there are these alternative pronunciations, and so once you decide on a phonetic transcription, you've decided on one. So there could be some consequences for this.

So number one, who does the transcription and who decides that's the one to go forward with?

DR. DORR: So there are two questions.

First, who does the transcription? I should clarify. These were all automatically transcribed, which means a choice was made and probably the wrong choice in many cases. One deterministic choice was made. So there was no human involved in that. On the basis of information on English in general, we know that -- and in fact, it probably would have come out with quinine. Who knows? But based on what it has available in general, we have an

automatic transcriber.

However, the second question is, what do we do with these different variants? What do we do with different pronunciations within a dialect? And then what do we do when you have different dialects entering into the picture? That's sort of the next phase of what we're trying to look at. We need to be able to train on different dialects in getting the variations of particularly vowel sounds. Those tend to be the ones that people trip up on the most. And even in different languages, which is another area that we want to look at next. Right now, there is just one deterministic answer and it could be the wrong one.

DR. DAY: Even within the same dialect -- in our lab, we have people just pronounce drug names and we find great variation even within very narrow sets of people, all highly educated, excellent readers, and so on. There are alternative pronunciations. Since what we're looking at is comparison of phonological similarity across pairs, if we don't have a sense of the alternative pronunciations and their relative probabilities of each one to begin with, then I don't know what we're comparing.

DR. DORR: No. That's exactly how you want to do it. You want to have differing probabilities with alternatives that are available to you, and what you rely

on is that if some vowel sound was wrong, that the remainder of the word would get you close enough that there's at least some hint that something could be going on here. But you do need to have more than one pronunciation, and as I mentioned, definitely within dialects, you do get these variations and people having the same education level will pronounce them differently. So I agree that that's something we are not doing now that needs to be done.

DR. DAY: Okay. And just my second question and last question. You've done a great job with the different features for producing the different sounds. There's often an interaction across features. So, say, for example, place and manner of articulation define stop consonants, and there's a huge psycholinguistic literature that shows that people make systematic errors in perceiving them. So these are sounds like "puh," "tuh," "kuh," "buh," "duh," "guh." And when people listen to those and make mistakes under noise or under good hearing conditions, you can predict what mistakes they're going to make. So they're more likely to confuse "puh" and "buh" than "puh" and "guh." These are direct calculations based on the number of features that vary.

So have you taken into account these well-known interactions of features in these computational linguistic methods?

DR. DORR: That's exactly what the decomposable features are supposed to give you, that you're not just taking "puh" as one sound, but you're breaking it down into, say, eight or nine different features. So that's where you can get that multiplicative interaction, that you have so many of them that it describes really a bunch of different dimensions along which you can compare another vector of features so that they differ in two of those features, but if seven out of the nine match, then that's a very highly likely confusable pair. And that's based on the phonetic literature.

DR. DAY: So how do you determine those weights? We saw 40 and 50 for place versus manner or vice versa.

DR. DORR: Right. That's tuning that was used initially for the cognate matching task for determining across language pairs like French and English whether there are certain similarities like couleur and colour, and those had to be retuned and adjusted so that, for example, manner and place are now given a higher weight than they were in the cognate matching task based on what we found in the data from the drug name pairs. So you can actually fine-tune it for your particular application.

As I said, the caveat is we were training on data that had other things playing into it that had nothing

to do with either look-alike or sound-alike names. A lot of these were reports and not real errors that actually occurred. So we were training on sort of noisy data, and we'd like to have a better training set to do that.

DR. GROSS: We have two more questioners and then we'll have to move on. Jeff Bloom.

MR. BLOOM: Yes. Dr. Dorr, can you come back up for just a second please? Thank you.

Picking up on what Dr. Day said -- and I would quibble a little bit with the vowel situation. We are living increasingly in a multi-cultural society, including not only just patients, but also doctors, nurses, health care practitioners, where there are particular diphthongs that are not native to their natural language, if English is not their first language. The R's and L's are particularly difficult for people to say. I don't know how that could be formulated in to figure out how to do that in what you're doing, but I think it's an important issue.

DR. DORR: And that's exactly what we're going to be doing next. We have a phonetic transcription table for Spanish, and we're looking at one for French. Again, these are superimposed on top of -- well, they're not really English names. They're some sort of brand name. So we're taking kind of what people would think a Spanish speaker would say an English pronunciation, and that is the

next phase. It's not a part of the work we've done so far. It's the next phase of the work. It's very important.

DR. GROSS: Stephanie Crawford.

DR. CRAWFORD: Dr. Dorr, please stay.

(Laughter.)

DR. CRAWFORD: I have two questions. First, when you were discussing the tests of orthographic distance and similarity, several times when you made the comparisons with Contac versus Zantac and Xanax, you stated it was not the result that you wanted to get. I'm a little confused with that because through objectivity, do you have presumed results you wish to get? That's the first question, and then I'll have a second one for you.

DR. DORR: First question. So we were again looking at a gold standard and did not find Contac and Zantac in there. Did anybody find that pair? If you did, let me know. If it shows up -- by the way, it will show up in the list. It will just be ranked lower, and so it depends where your threshold is. But Xanax and Zantac is a confusable pair and Contac and Zantac were not among the confusable pairs. The reported pairs. So that's what I'm saying. It seems that that's the result we wouldn't want.

DR. CRAWFORD: And my last question. I appreciate the very fine comparisons you did with the three approaches, ALINE, Dice, and LCSR. I wanted to ask, are

these the only approaches? If not, how were they the three that you selected for comparisons, and if they're the only ones you're considering.

DR. DORR: So LCSR and Levenshtein are actually related. There are also other versions. Like there are bigram and trigram versions of these. I put the sort of simplest cases up there, but we did take the string matching approaches that were in the computational linguistics literature to be reported the best in our comparison. And then phonological -- the standard -- when we began studying this with Soundex or its sort of relative Phonex which we also looked at. We just started with what was reported to be best in the literature for each of these types.

DR. GROSS: Thank you all for those excellent questions.

The next speaker is Dr. Richard Shangraw, Jr. who is CEO of Project Performance Corporation. He will discuss the use of expert panels in evaluating drug name confusion.

DR. SHANGRAW: You can tell already we're going to change gears a little bit here. My presentation is sort of at the other end of the spectrum. It's really talking about the use of expert panels as a way of identifying potentially confusing drug name pairs. In some respects,

it's going to build on Bob Lee's comments about the use of experts in this problem area. And I'm going to talk a little bit more broadly about the problem. In fact, when I got the questions for this presentation, I interpreted the question about how does this method compare to others to be a broader question about how does expert panels, for example, compare to computational linguistic approaches or experimental pharmaceutical approaches. So I'm going to have a sort of broader perspective on the problem.

Before I get into the problem set, let me just give a quick background for those who may not know a lot about the field of expert panels or expert committees. It's an area that has emerged primarily in the '40s and '50s. It grew out of a lot of research on the use of experts in a number of different settings: policy settings where there were some concerns that policy makers here in D.C. were not generating the best policy decisions when they got together to solve problems. That led to a number of formal structure techniques for using expert opinion. I don't think they use them now, but at least there were some thoughts of trying get those structured techniques in place.

You'll hear through my presentation today the use of the term Delphi. There's a technique called Delphi that's been used as a nominal group technique that's been

used formally for many years, 20-30 years.

And there's also been a large application of the use of expert panels in the health care field. In fact, there's a longstanding set of research that's been done by UCLA and the RAND Corporation on using these kinds of expert panels and approaches for looking at appropriate care in hospital settings. NIH uses a consensus-based approach for some of their decision-making.

I think Dr. Seligman was accurate in saying that there hasn't been a lot of specific research in this problem set area, that is, the use of expert panels in this drug name confusion area, but there's a load of evidence and research in using expert panels in many other settings. What you're going to hear today is my bringing that amount of expertise and that research that's been done into this problem set area and talking about a process for how it might be used for drug name comparison purposes.

I'm going to be very procedurally oriented today. I think the biggest criticism of expert panels and expert committees is the ability to replicate or validate their outcomes. The best improvement that can be made in terms of improving the outcome of an expert panel or an expert committee is by introducing repeatable processes related to the way that these panels or committees are conducted. As you'll see here on my slide -- and this is

really going to be the driver behind my presentation here -- I'm going to work through a design on how an expert panel could be conducted that could be replicated and perhaps validated -- and I'll talk about them a little bit later in the presentation -- as a way to ensure that you could get consistent and possibly highly appropriate results coming out of a group of human experts as opposed to a computational system on a computer.

I'm going to go through each one of these boxes, but in broad terms, there is a panel that's selected and moderated, and before you can really select and moderate that panel, you have to figure out the definition of who's an expert and you have to figure out what sort of guidelines this panel is going to use in terms of the way they vote or rank decisions through the panel.

Most of the literature talks about and most of the research that we've done talks about the use of separating these panels into rounds or phases where you would have the problem set introduced. It's often called the exploratory round or the discovery round where you actually try to just put on the table all the possible alternatives where you might have a confusion with a specific drug name. You would then consolidate and collate those results.

Then you would have a second round where you

would have a ranking or voting process. In fact, some of the techniques I described earlier, the nominal group technique and Delphi technique, will extend these rounds many times. They'll go three rounds, four rounds, five rounds before they come to an actual decision.

Then obviously you'd have some solution set or result coming out of this panel.

Perhaps the first problem and probably one that's most challenging here is to make sure you have the right experts participating in the panel. Again, guidelines can be established here. It can be based upon experience. It can be based upon not only years of experience but type of experience, clinical experience. It can be based on education, training, pharmacists, nurses, doctors. But clearly there could be some baseline established here for the type of expert that would be asked to participate in the panel.

Second, you have to be concerned about conflicts. This is an interesting problem that you've already discussed this morning in terms of this panel being put together in terms of making decisions. This is clearly an expert panel sitting before us here, and you have to be concerned about those in these kinds of panels also.

Personalities is a clear factor of concern that's been introduced through many studies. The concern

here is on dominating personalities. Obviously, in the front-end stage, you certainly don't want to select a whole set of dominating personalities to be part of your panel.

Then finally, there's some good research now to suggest that the larger the diversity of the panel, the more likely you are to get a broader or more robust result. So, in other words, if the set is all pharmacists, it's probably not as good as a set that has some pharmacists, some nurses, some doctors. You even heard Bob Lee talk about the fact that they introduce legal counsel into their panels and other people that have expertise in this area.

The second part, again before you even get started, is laying the groundwork on how you vote and how you rank decisions. This is another very important part of the process. This is probably the part of the process that can lead to the most dynamic changes in the outcomes of panels. These are very simple issues. Does the majority vote win? If you pull a pair up and the expert panel looks at it and the majority thinks it's a problem, is that sufficient? If it's not majority, is it two-thirds? If it's not two-thirds, is it 90 percent? Making those decisions on the front end before you get to the process, obviously makes a process more repeatable.

And the second part of that is related to how you collate the results. If we have 10 experts in a room

and they're trying to vote on or rank a set of problems associated with a confusing drug name pair, how do you rank or collate the different ranks amongst the experts? There are a number of different techniques out there for doing this. The nominal group technique has an extended process that looks at the way that people rank and combines those ranks together, giving higher priority to first and second ranks. We could spend a long time talking about just how you collate ranks, but suffice it to say there's a process for doing that. There are different ways of doing that. None of them are perfect, but at least you need to establish that on the front end.

You've seen some numbers already today from Jerry Phillips about numbers of participants in their expert panels. I think you'll hear some from some of the other speakers today. Dr. Kimel, for example, who's up after me, has a very closely related area and that's use of focus groups, and she'll talk about some of those numbers also. But in general, the size of an expert panel is about 8 to 12 participants.

The issue of moderator, which I'm actually not going to spend a lot of time on because Dr. Kimel is going to spend some time on it, talking about the role of the moderator. It's also very important in these groups as a way of facilitating the discussion.

So now let's break it down into how an expert panel would proceed. Round one. Given the electronic age, most of the expert panels that we're seeing being conducted out there are certainly from a cost perspective in terms of making sure they minimize the cost of conducting these panels are conducting round one's electronically. It's predominantly done through e-mail. An e-mail is sent to a participant. They are given some procedures and processes about how they're to look at different drug names. They're asked to provide a ranked list back to the moderator, and then those ranked lists are collated. Clearly the number of names being processed by an individual, the ranking procedure and process can all affect this stage of the process.

There are also clearly some concerns here given this topical area of confidentiality. I'll talk about that a little bit later in terms of strengths and weaknesses of this approach.

Once you get the results for round one, you consolidate them, using any of a number of different approaches for taking ranked results and putting them together and displaying them. Some of those approaches simply say let's just focus on the number one rankings from across the experts, and there are also ways of taking those rankings and consolidating them in such a way that you can

have a broader list exposed to the participants in round two or a narrower list.

Again, this is an area that Dr. Dorr hit on just briefly, and that is the issue of if the system, whether it's an expert panel or a computer system, generates potentially confusing names of 100 potential pairs, it's much more difficult to rank in order and organize those types of results than ones where you see 10 or 20 potentially confusing names. This process, while it seems much more human based on the computational methods, can yield the same kind of results where you could have potentially very large sets of potential confusing names coming out of the set of experts, and you have to be concerned about the ability of the experts to process through those names.

Round two is really probably the round that is the focus of most of the expert committee/expert panel research and that's really the way that you get at the decisions. It's called the decision round, summary round, the ranking round. It's the part in the round that after the discovery round, round one, that you bring the experts back together and have them now, in a face-to-face situation or increasingly in a computer-facilitated situation, discuss the potential issues associated with name pairs or potentially confusing name pairs.

As I said before, this is a process that historically has been done face to face. Experts are flown in, for example, this panel you see before you here. And they are asked to communicate amongst themselves with a moderator, to sort through a set of issues. Increasingly there are web-based tools that are doing this where you have a speaker phone, a teleconference, augmented by a computer screen on the internet where they're able to have conversations through the telephone lines, and they use the computer screen as a way of organizing and facilitating the discussion.

Again, there have to be some predetermined rules about voting. This process can be a lengthy process. it can take anywhere from 2 hours to 6 hours to 8 hours depending upon the complexity of the name that's involved. It's also an expensive part of this piece of this process given especially the cost, for example, of flying this group of experts in. You can imagine the cost of doing that across the 300 or 400 names, for example, that Jerry Phillips says has to be reviewed on an annual basis.

So can we validate these methods? Obviously, the biggest concern here is can you replicate across expert panels the results of the expert panel. Most of us sitting around here today would say that's a tough problem. Right? Experts have different perspectives. They come from

different views. They're moderated differently.

I would argue that if the procedures and processes are well established ahead of time and if there's understanding of those processes by the participants, if you have diversity of views, and you have a good moderator, that there is a possibility of replicating these procedures. It could be done two ways from a testing perspective.

The first is one which I call reliability. That is, do different panels come up with the same results? That's the first question. So if I have one panel here today and a panel tomorrow and I give the same drug name, will they basically come up with the same result? Obviously, that could be tested. It's expensive to pull those panels together, but nevertheless, it could be tested.

Second is the issue of validity or in this case predictability, and that is, if the panel is given a name, do they come up with an answer or a potentially confusing pair that can be compared against some standard? We've talked about this gold standard in the first talk by Dr. Dorr. That again could be replicated giving a panel a set of names that we know have known confusions on and see if they actually generate that same list of names whether there are known confusions. Again, that could be tested.

It's expensive, but it can be done.

There are some problems, of course, in that second test in terms of what's called the history effect, and that is, if panel members know that there have been known confusions with a name, then we have problems in terms of history, with that effect. But nevertheless, you could perhaps control for that in terms of panel participation.

So these are probably the two key pieces that you'd like to look at from an expert panel perspective.

So what are the strengths of the design? Well, clearly when Dr. Dorr was asked the question by one of the experts here on the panel is this approach sufficient in and of itself, and that was asked on the computational approaches, I think much the same question could be asked about an expert panel. Is an expert panel sufficient in and of itself to solve this problem or to address this problem?

And my answer, being a good social scientist, is that I'd always like to have multiple methods. So a combination of a method, for example, of a computational approach perhaps on the front end for the discovery phase, which is to say, give me the list of potential confusions, and then taking that list and providing it to an expert panel, much like the process that Jerry Phillips describes

the way that the FDA does it, seems to me to be a more appropriate and possibly more robust approach to solving the problem because in my opinion the ability of the human expert to digest and to analyze some of the questions that have already been presented by this panel, in terms of the computational approach, could have some value, the ability to sort through dialect by different pronunciations, by misinterpretations, by handwriting. These are all things that the computer is getting pretty good at, but I still think the human has an ability to do some more in that area.

Second, I think the other part of this, which is the really interesting piece of this puzzle and that is with a set of experts sitting around a panel talking about potentially confusing pairs, you can ask the panel why do you think that's a confusion. It's hard to do that with a computer. In other words, you can say why is that confusing to you, and you can at least get some elicitation from the expert about why they think there might be a confusion. Now, we could probably dive into the mechanics of why the computer thought it was a confusion, but I think as a group of reasoned experts in a room, you like to hear a human interpretation of that potential confusion.

And finally, as you can see, the design is easy to understand. It's pretty straightforward. It has some

process pieces to it, but it's relatively easy to understand.

Weaknesses. Many weaknesses with this approach.

I talked, first of all, about the fact that the panels are susceptible to domineering personalities. We've already talked about that.

It's difficult to validate the designs. I proposed some methods, but they are difficult and require a lot of controls.

The ability of the group to achieve consensus is a particularly perplexing problem with expert panels, in that even if you establish voting methods, there may be some issues in terms of the ability of the panel to come to some sort of consensus-based conclusion.

We've already talked and heard some issues about dialect and concern. If the panel is not diverse enough, there may be some issues there.

You can also have wide variability in the results across panels given the expertise of the panels.

And finally and probably as important is as we move to these electronic panels, there's always going to be concern of confidentiality, certainly on the part of the pharmaceutical industry in terms of taking these names and putting them across the ether to other people to comment on

them.

So that's a quick overview of the expert panel and expert committee approach to this problem.

DR. GROSS: Thank you very much, Dr. Shangraw.

Any questions from the advisory committee? Yes, Eric Holmboe.

DR. HOLMBOE: I'd just be curious to know, with regard to expert panels, what data do we have with regard to this issue in the past? You mentioned, Jerry, that about a third of names get rejected. What role have expert panels, if any, played in that particular process?

MR. PHILLIPS: The expert panel plays an important role in our process, but it's just one component of a multi-faceted review. So I think if we went back and looked at the recommendations of the expert panel on the final conclusion, that they're going to be pretty consistent.

DR. GROSS: Stephanie Crawford, do you have a question?

DR. CRAWFORD: Thank you. A very quick question. How do you determine consensus? You said it's not always achievable? By what definition would you have consensus?

DR. SHANGRAW: Well, the first problem with consensus is and the failing of many of these panels is

they don't decide on the voting method before they have the panel. So if you don't decide on the voting method before you conduct the panel, you will never get consensus. Certainly it's harder to achieve. So the first solution to that is to have an agreed-upon voting method before you go into the panel process.

Voting methods can be as simple or as complex as you want them to be. Some use simple one vote mechanisms. Some use majority mechanisms, plurality mechanisms. Some use rolling voting mechanisms. There are a number different techniques. But the most important point here is establishing that ahead of time and having the panel participants agree on that. If you do that, then consensus is easier to accomplish, obviously, because once you get to that point, you hold the vote, and whatever voting method you've decided to use then helps to finalize your consensus.

Unfortunately, most panel members, after a long and heated debate, when they get to the point where they're supposed to vote, decide they don't like the voting methods. And then we have another set of problems. But that's the difference in dealing with humans than with computers.

DR. GROSS: Michael Cohen.

DR. COHEN: Yes. It hasn't been mentioned yet,

but I think a large percentage of the practitioner review that Mr. Lee was talking about before is actually done by companies that are separate from the PhRMA company. I think most of those companies, from what I can gather, use a system where they would actually -- first of all, there's more than just one name that's tested for a particular compound. There might be 10 or even 15 or more. But they would use what is considered, I think, an expert group. In other words, there are physicians, nurses, pharmacists that are out there in the field that are working every day, and it might be done by the internet. They would actually look at actual names and listen to them, how they're pronounced, et cetera, whatever, and then provide feedback. And then that information is collated and presented to an expert group that does what is called the failure mode and effects analysis or failure analysis.

Is that considered expert panel on both ends? That is not.

DR. SHANGRAW: No, absolutely.

DR. COHEN: Oh, it is.

DR. SHANGRAW: You're going to hear from the next speaker an even broader discussion on focus groups, and we can have a long debate about is an expert panel the same as a focus group. The answer is they all come from the same genre. They all come from the same category of

approaches that says let's convene a group of human experts. Let's tap into their brains and let's find solutions to problems. So the next speaker is going to talk about that from a focus group perspective, which in fact some of the third party research groups use focus groups, and she'll be talking more about that.

DR. GROSS: Brian Strom has the next question.

DR. STROM: We've heard today, it sounds like, an enormous effort underway at FDA and industry, multiple private companies using expert panels. This has been underway for many years, it sounds like. You described for us a very clear, very nice description of the process and how you would test the reliability and validity. Given the huge effort that has been underway all these years, all these drug names, can you tell me what data are available on the reliability and validity of the approach?

DR. SHANGRAW: If the question is what's available on the reliability of an approach testing drug names specifically, I do not have any data in that area. That's not to say there's none out there. I'm not aware of any at this point.

DR. STROM: Does anybody know? Jerry?

MR. PHILLIPS: I'm not aware of any either.

DR. COHEN: I don't think there is any.

DR. SHANGRAW: It's sad that we don't because

you're exactly right. We've been doing this for years and we should have some data, but I haven't seen it yet.

DR. GROSS: Jeff Bloom.

MR. BLOOM: Thank you.

In reading through the meeting materials and also your presentation, one of the things I was wondering about is, has there been any consideration of including patients in any of the expert panels? After all, patients need to understand the drug names and also serve as a check and balance against making sure they're getting the correct drug.

DR. SHANGRAW: In many of the health-related expert panels, for example, ones convened by NIH and UCLA, there is a role for the patient in those panels. Obviously that comes into the front part of this discussion where I talked about how you define an expert, and clearly that would be part of that discussion about whether or not a patient would be included. I think there are a number of reasons why you might want to include a patient, but that would have to be determined on the front end.

DR. GROSS: There's a question or a comment from Jerry Phillips.

MR. PHILLIPS: Rick, the process in which you vote in an open meeting, whether that's privately -- what influence does that have on the decision-making process and

how important is that?

DR. SHANGRAW: That's a very good question, and I failed to address that. One of the techniques that has been used to deal with the domineering personality problem in expert panels is to use anonymous voting throughout the process. Now, there's been some research on that which says that a completely anonymous voting process, especially in the expert panel, that second phase, which is the decision phase, doesn't lead to the best decision because you have to expose at some point a position and then use that as a basis for discussing the problem. So the general approach has been, in the literature at least at this point and the research, is to have anonymous voting through phase one, which you saw in this process, which is to identify and rank on an anonymous basis through that discovery phase to present the list, but then by phase two, that that voting would become more public as a means of facilitating discussion. There's a longstanding debate about even if you have that open voting process in that second phase, and there are still some that argue to keep it anonymous, but that it is a key piece of the issue of the domineering personality problem.

DR. GROSS: We will adjourn and reconvene at 10:35.

I have a suggestion for FDA and PhRMA. Maybe

at lunchtime you could prepare a list of what methods you are currently using to avoid look-alike/sound-alike names so that when it's time for us to make some recommendations, we have that information summarized for us.

(Recess.)

DR. GROSS: I hope you all had a nice coffee break. We're going to reconvene so we can try to stay on schedule.

The next speaker is Miriam Bar-Din Kimel, Senior Project Manager of MEDTAP International, who will talk on the focus group methodology.

DR. KIMEL: My presentation will be about focus group methodology and the application to the drug naming process. It will actually build upon similar methods that Dr. Shangraw had discussed in the previous session.

First I will review focus group methodology, including strengths and limitations. Then I will describe how focus group methodology may be applied to the drug naming process, and finally discuss conclusions.

Focus groups are a form of qualitative research methodology used to address specific research questions that require depth of understanding that cannot be achieved through quantitative methods. Focus groups can be used in various phases of research and in conjunction with various research methods. In the exploratory phase, they can help

determine which populations to test and to target. In pretesting, they can help identify and clarify perceptions about specific topics, products, or messages. And in triangulation, also known as convergence of multiple data sources or methodologies, focus groups can be used to support other sources of qualitative data.

More specifically, focus groups can be used to gather background information, diagnose problems with programs and processes, stimulate new ideas or identify new relationships, generate hypotheses for future qualitative or quantitative study, evaluate programs, develop qualitative understanding of how individuals view a situation or deal with a phenomenon of interest, or help interpret quantitative results.

Focus group methodology can be used as a standalone investigation or as part of a multi-method study in conjunction with other qualitative and quantitative methods. For example, in survey design, focus groups are often used as a first step to identify relevant items in the patient's own words. Once the instrument is developed, quantitative psychometric analysis is then performed to test the instrument properties.

Focus group methodology also can be used to supplement the interpretation of quantitative data. For example, a trial may find a large number of asthma patients

come to the ER for treatment with minor symptoms, and then a focus group can be conducted afterwards to find out why they come to the ER.

There are different types of focus groups that may be used. Traditional focus groups are conducted in person and have a structured format, most often using interview guides to direct the discussion. Brainstorming is also conducted in person but is nondirective and unstructured. Delphi techniques, as previously described, can be done via mail using structured questionnaires to direct participants to identify issues relevant to the topic of interest and then rank the issues in order of importance.

Traditional focus groups typically involve 8 to 12 individuals who discuss the topic of interest under the direction of a trained moderator. The moderator must be trained in group dynamics and have strong interviewing skills. This is important to avoid domination of aggressive individuals in the group and to include quiet individuals. They are structured and use an interview guide to help direct the discussion. They last from 1 to 2 hours depending on the research question and the characteristics of the participants. A recorder is generally used to take field notes during the session. Findings are often transcribed from the recording.

For in-person groups, facilities designed for group interviewing are ideal, enabling members of the scientific team to observe the discussion and, if consistent with the study design, provide the moderator with additional questions or queries pursuant to the group's discussion.

Focus group participants are chosen based on characteristics that the researcher wants to understand further, also known as break characteristics and control characteristics. The number and nature of the groups and sessions is determined by the purpose of the study, the design complexity. For example, if the characteristic of interest is complex, a researcher may want to conduct several focus groups to make sure all relevant themes are identified. But typically two to three focus groups are conducted in diverse geographic regions, and the nature and number of groups is also based on the resources allocated.

Data from focus group include tape recordings, transcriptions, which for a 2-hour session could be up to 40 to 50 pages, and field notes which are usually taken by a second researcher during the focus group session.

The analysis is driven by the underlying research question and involves a careful review, synthesis, and summary of data from tape recordings, transcription, and field notes. Qualitative data is interpretive and

constrained by the context. In addition, the topics are generally linked to the interview guidelines. Data gathered during the focus groups take the form of information, quotations, themes, and issues gathered from the participants during the course of the interview.

Steps involved in data analysis are mechanical, such as organizing, and interpretative, such as identifying common themes and patterns within themes and drawing meaningful conclusions. Software such as Ethnograph may be used to help identify themes.

Reliability of data may be enhanced by repeated review of the data and by independent analysis by two or more experienced analysts.

Results are expressed qualitatively as themes, issues, or concerns and are highlighted with substantiating quotes. Results also may be presented quantitatively such as the number of participants who agreed or disagreed on particular issues and the frequency of themes within the group discussion. The appropriate sample characteristics are also presented so the reader or the reviewer has an understanding of the nature of the participants providing the data.

Focus group methodology is only as useful and as strong as its link to the underlying research question and the rigor with which it is applied.

Strengths of focus groups are: that they provide concentrated amounts of rich data in the participants' own words on precisely the topic of interest; that the interaction with respondents and interaction among group members add a richness to the data that can be missed in individual interviews; and that the data can provide critical information in the development of hypotheses or the interpretation of quantitative data.

The primary limitation of focus group methodology is the relatively small number of participants and the limited generalizability to the larger population.

Group dynamics can also be a challenge or a limitation. A group with particularly quiet individuals or aggressive talkers or a group with a tendency toward conformity or polarization can make group dynamics difficult, particularly if the moderator is inexperienced. Careful attention to study design replication using multiple groups within a study and a well-trained experienced moderator can minimize this limitation.

In some cases, interpretation can be time-consuming and require several experienced analysts. To enhance the strength of the results, independent analysis by two or more analysts is always preferred.

Focus groups may be a useful method for identifying problem areas in testing proprietary drug names

to minimize medication errors. For example, this methodology is ideal for understanding potential sources of confusion from the user's perspective, and therefore focus group participants include physicians, pharmacists, and nurses, as well as patients and caregivers.

Focus group methodology also can be used to identify situations in which confusion is most likely to occur. For example, in particular patient populations, such as elderly patients taking multiple medications or situations such as pharmacies where drugs are shelved alphabetically by proprietary name.

Focus groups can also be used to test conclusions of expert panels about sound-alike medications that pose a threat in the practice or home setting, to develop research methods for testing sound-alike medications quantitatively, and for understanding behaviors underlying prescription practices that can contribute to name-related errors in order to identify high-risk therapeutic areas.

Focus groups can also inform quantitative research design; provide qualitative data to aid in the interpretation of quantitative results, for example, explain unexpected areas of confusion; serve as an integral part of a multi-method evaluation program, for example, triangulation with in-depth interviews with physicians,

pharmacists, or patients; and provide a useful foundation for designing risk assessment and management studies, for example, identifying potential problems in professional practice and home use patterns.

When used appropriately, focus group methodology can provide rich depth of understanding of a problem or phenomenon of interest. Depending on the response question, it can be used in isolation or to complement or supplement quantitative methods. And as is true of all research methodologies, its utility is a function of its link to the research question and the rigor to which it is applied.

DR. GROSS: Thank you, Dr. Kimel.

Any questions for Dr. Kimel? Yes, Lou Morris.

DR. MORRIS: In your conclusion, you say it can be used in isolation, but in all the examples you gave, it seemed to be used in combination. Could you describe a situation where you think it could be used in isolation?

DR. KIMEL: In general, I think it could. Probably for the purposes of working with drug naming, I think it would probably be best to be used in combination.

DR. GROSS: Any other questions from the panel?

(No response.)

DR. GROSS: If not, we'll move on to the next speaker. Kraig Schell, Assistant Professor, Department of

Psychology at Angelo State University, will discuss use of laboratory and other simulations in assessing drug name confusion.

DR. SCHELL: Good morning. Let me start with a couple of preliminary remarks: first, to tell you what a privilege it is to be here with you this morning, and second, to express deep regret that unfortunately Tony Grasha, whom many of you know, who would have been here today, of course passed away about a month ago. So I'm going to do my best to fill his very, very large shoes. A lot of what I'm going to talk about today was research that he and I had worked on for now the past seven years. But, unfortunately, a good part of it is also in his head, and so I'm going to do the best job I can to try and estimate what would have been in his head with respect to some of these topics.

The current state of the problem, as we've seen it, he and I, over the past seven years, is clearly that drug name confusion is a component that we need to be concerned about with respect to patient injury and financial loss. Many of the means of assessing drug name confusion are primarily based on rational and reductionistic approaches, such as FMEA and RCA, phonological and orthographical analysis and expert teams and committees, which all three, to some extent, are based

on a rational decision-making approach to the problem. Unfortunately, as we know in psychology for quite some time, humans aren't necessarily rational. In fact, we're rather irrational things, and the problem of name confusability is also a broad and less rational problem than might be assumed just by looking at it superficially.

Some of the research that we've done over the last seven years has identified many of these factors, as well as several others that I didn't have the room to list, as potential problematic variables that can affect error production and error capture in pharmacy filling and verification tasks done both in our laboratory at the University of Cincinnati and also at Angelo State University where I am and also in field sites that we've worked with over the past few years.

Our approach to the problem is based on these following assumptions and observations. Drugs that look and sound similar are not confused with each other or misfilled, at least with the current data we have available, in the same proportions that we would expect based on their similarity indices. For instance, Zantac and Xanax which was talked about before. Obviously very similar phonetically and also has quite a bit of similarity in terms of its bigrams and trigrams, but you would expect that with degree of similarity that we would be misfilling

that drug 7-8 times out of 10. Thank God, that's not the case. Actually we're much more accurate than that.

That leads me to believe that that variable, although it is important phonologically and orthographically, is not the only problem obviously. And I agree with what Bob Lee said earlier. There are definitely other conditions that need to be included and added into the equation such that perceptual factors are necessary, but not necessary and sufficient explanations for why the problem of human error exists.

And the third assumption that we rest on is that human error as a process is not rational. In fact, Dr. Riesen, in his classic work in 1990 on human error, called errors latent pathogens that sit inside systems and processes in every organization and every realm of society that are just waiting for a situation to bring them to the surface and infect it with an error.

I'm reminded of the problem that occurred with the USS Vincenz and the Iranian airliner a few years ago in the Persian Gulf, and if you evaluate that particular topic very closely -- and many people have in the psychological literature -- you see that the individual components of that particular event weren't necessarily problematic in and of themselves. It was the combination of those components in that particular given situation that led to

the erroneous decision to shoot down that airliner. That's the approach that we're taking, which is much more consistent with a human factors approach to the problem, much broader in its scope.

So simulating, as we do in the research we've done for the past seven years, gives us the ability to look at human factors that might interact with the physical characteristics of a drug name. In other words, under what conditions are Zantac and Xanax more or less confusable?

One possible thing that we could talk about here -- and I'll mention it again later in the talk -- is the informational context surrounding the drug. For instance, Mr. Phillips talked a little bit about the Avandia and the Coumadin misfill and mentioned in his talk a very important point, that the dosage and the administration of the drug is probably a significant contributing factor to the confusion of Avandia and Coumadin, two words that look, as he said, relatively nothing alike. And it's those kinds of factors and those kinds of issues that we can look at in a simulation paradigm.

This is the model that we are proposing that Dr. Grasha and I built and I am proposing it to you today that the simulation structure should take. Along the left-hand side of the slide there, you see what is called the

control/realism continuum. Generally speaking, as control increases -- in other words, as experimental control is strengthened -- the realism of the simulation decreases. So the stuff at the top of the pyramid that you see, the lab simulation and the pharmacy school simulations, because of the necessity of experimental control in those paradigms, they're necessarily going to be somewhat artificial and they're going to eliminate sources of variance that could be important.

As you progress down the pyramid to the error monitoring stations, there we have a great deal more realism as we're actually working in pharmacies and hospitals around the country, but the control that we have over error production and error capture is lessened. It requires the complete model to get a full and total picture of how medication errors exist and are produced and are captured. Just looking at one of these levels is not going to give us a complete picture.

The simulation also allows us to capture what we call a subjective error. Basically what that is is an error that is made and is corrected before it leaves the pharmacy. These are a significant source of error in our research that are not going to be predominantly recorded in self-reporting databases such as USP, et cetera. The objective error would be the error that actually left the

pharmacy and then was recorded as one of those that occurred. We call them also process errors because they are errors of the human process that is required in order to fill or verify a script from beginning to end.

One very interesting finding that we replicated numerous times in both the laboratory and in retail and outpatient pharmacies is that for every six process errors that we can capture, one of those tends to get by all verification steps and actually leave the pharmacy and be dispensed to consumers. We believe that's a very important ratio because if we can demonstrate that a particular drug is creating an inordinate amount of process errors, that gives us pause and makes us begin to think that if that drug name were allowed to be put into actual pharmacies, running the risk of pharmacists being more vulnerable to moving into an error mode of processing and then, as a result, more of these scripts actually leaving the pharmacy.

Another benefit to the simulation is that it's safe. None of these drugs actually go to anyone and they aren't actually taken by anyone during the simulation. So we can make as many errors as we want to and no one is actually harmed by them. In fact, one of the designs that Tony was going to do before his untimely passing that we talked about for several years was to use the simulation to

train for errors, as has been done in other fields, where we actually force the participant in the simulation to make the mistake over and over and over again, to build a schema for that mistake so when they do it later, they recognize, wait a minute, it's not right, I shouldn't be doing this, this doesn't feel correct, and they're able to make a correction.

It allows us to use a variety of different experimental and quasi-experimental designs. We can do case studies. If we wanted to look at team performance in the pharmacy techs and pharmacists and how they're interacting, we can do that. We can do an actualistic observation design, a variety of different approaches are possible in the simulation.

And we can insert drug names that are being evaluated into an existing database of already evaluated and marketed drugs to see if anything currently on the market that maybe we haven't pinpointed up to this point is a source of potential error that we may have overlooked.

Three laboratory approaches that I can talk to you about. Two of them we've done already. The third one is in production right now.

The full-scale dispensing task is exactly what it sounds like. We use mock materials to allow participants to fill mock orders for these prescriptions.

It's actually rather amusing if you were to take a look at it, and I'll show you a picture in a moment. We use things like craft beads and paper clips. We even used cereal at one point in time. We had some Trix cereal on the shelf that we were calling drugs and assigning names to them and having people dispense them as if they were sitting in front of a bench in a pharmacy.

The verification task is where the scripts are filled beforehand and an individual takes the scripts in sets, verifies them against a database with the same information that would have been on the label, and then tells us whether this order is correct or this order is not correct. Very similar to what a pharmacist might do going back to through the will-call or the return-to-stock bins to see if anything was erroneous in that sense.

And thirdly, the drug name perception task following the methods of Bruce Lambert and also Dr. Dorr, what she's doing. I'm building this currently at Angelo State University to be able to look at drug name confusion from that human factors perspective, being able to add different individual difference factors and see how that influences the confusability of the names.

That's a panoramic view of the original pharmacy simulation lab. It didn't reproduce very well in your handout, but essentially it's just portable plastic

shelves with a computer work station. The scripts were written on index cards and in various styles of handwriting, and participants simply sat in front of the computer and were able to fill the scripts as if they were working in a pharmacy. They do sit. Pharmacists for the last few years have told us how unfair that is because they always have to stand.

(Laughter.)

DR. SCHELL: The only explanation I can offer you is we didn't have any tables that were tall enough, so we had to make do with what we had.

This is the verification lab I currently run at Angelo State University. On the right-hand side, those are the scripts. We use standard 30-count pill bottles. You'll notice that there is a 3-by-5 index card in each of the bags. We use that to simulate the label that would normally be attached to the bottle, and we chose to do that primarily for convenience. The labels would eventually tear or start to lose their adhesion, and it would become an issue of cost. The index cards are much more durable, so it allows us to keep our costs down.

But the individuals simply look at each script, decide whether the correct item is in the bottle, whether the correct amount of that item is in the bottle, and whether the index card information matches a database that

they are presented with for that particular script.

The drug name confusion task. The interface for this is currently being built, so I'll describe it the best I can. Essentially a drug name would be presented to a participant on the screen and we'll be able to vary the amount of time they'll be able to see that name. Then they have to navigate through a virtual shelf where they have to select first what letter did that name start with. Then that will move them to a new screen where there will be a variety of different drug names starting with that letter, and then they have to select the drug name that they believe they saw.

Now, here's the kicker. Once they select one of the letters, they can't go back. So if they select a P, for instance, and then they realize, oh, man, it didn't start with a P, well, they're kind of stuck now. They're going to have to select the one that they think is closest to what they saw, realizing they've already made the error. The reason we make it so that it does that is so that we can separate process errors from committed errors. When each of those occurs, we'll be able to separate them out.

We can change the duration of name presentation, the inclusion of informational context. We can add feedback to tell the performer whether they're doing well or whether they're doing poorly at given

intervals.

The informational context variable I should also mention can be switched to a different domain of knowledge. Since we're looking at basic human performance and we're using primarily naive participants, most of our participants don't know quinine from Celexa. So dosage and administration information is relatively meaningless to them. So we have four different knowledge bases that are more in a college students domain, such as television, movies, sports, and things like that, and then we can provide informational context around those and study basically the same perceptual processes.

The pros. Strict control is the biggest advantage to the laboratory simulation. We can tailor that as necessary. We can vary systematically different factors that we believe to be important. What I mean by customizable products is that we can do more than one product name at a time. We can insert 20 different product names into a given experimental design if we wanted to, and provided folks are on task long enough, we could look at a variety of different permutations and combinations of those.

The disadvantages. The lack of realism.

Shorter versions of the task tend to be overly simplistic, and what I mean by that is the shorter that

they're on task -- and believe me, getting a college student to do anything for 2 hours is a chore. We have to eliminate a lot of things that pharmacists do such as take phone calls, be interrupted by customers, have to deal with insurance companies, and those kinds of things. Longer periods of time on task, we can add those things in.

It's possible that we might control some causes of name confusion and other sources of error in the experimental design per se. So numerous experimental designs and numerous studies would have to be employed.

The movie set simulation, the second tier, is a broader-based pharmacy simulation where the environment is more similar and more exact with respect to an actual pharmacy. The emphasis would be on duplicating the work flow and other conditions under which prescription filling and checking would occur, such as the insurance companies and the multiple scripts at one time, and the irate customers, and those kinds of things. Both objective and subjective data could be collected in this as well.

A note of explanation here. By training I am a business psychologist, and one of the things that many corporations do to select managers is something called an assessment center -- maybe some of you are familiar with that -- where management trainees will be placed in an observation tank, basically a large area, and given a set

of exercises to do while current managers watch and rate them. In the movie set simulation, we apply the same basic analogous idea to this particular level of the pyramid. We would be able to create exercises that incorporate many of these factors that could impact performance into a series of exercises that then we could do with each of these drug names.

So there could be the insurance fiasco exercise, for instance. How does dealing with an insurance company while you're filling a script for that particular drug name impact its confusability?

The multiple script exercise.

Similar preceding name. Much of what we've done to this point has been on looking at pairs of names simultaneously. Well, what happens when we have a consistent, frequent representation of one name, followed by then a highly confusable name right after that? Is there a perceptual bias toward the name that had been perceived first?

Frequent prescription exercise.

Stressed out exercise.

All these things that you see here could be designed and we could, just like the gauntlet, run a name through a series of these exercises to see how different environmental conditions affect their confusability.

The simulations in the colleges of pharmacy are very similar to the movie set simulation, but with one important difference. In the movie set simulation, the emphasis is on researching and pinpointing environmental and individual difference factors that could impact confusability. In the college of pharmacy, we would then take that knowledge into a similar situation in the college of pharmacy and then train new pharmacists on those situations in individual difference factors, being aware of them, understanding that they occur, understanding how they influence confusability, and be able to dedicate a little bit more training toward the confusability factors that enter into doing their job on a daily basis.

So in the movie set simulation, really basic research is the emphasis. In the college of pharmacy simulation, training is the emphasis. As a result, it may not be quite as flexible for manipulation and experimentation since training is a little bit different approach than basic research.

Finally, the error monitoring station. In automated pharmacies, especially the pharmacist's role is switched from filling to verification largely. As you, I'm sure, are aware, in many States now technicians can do most of the filling tasks by themselves. In Texas I believe a technician can do everything from start to finish. The

only thing that's required is that a pharmacist check the script before it leaves. So that's starting to become a trend. So verification is becoming more and more important.

This test would insert the new drug into an existing pharmacy that would be, of course, in connection with FDA or the pharmaceutical companies. Controls would be in place to ensure that the drug is not actually dispensed, but we would insert mock orders for this drug into the standard flow of everyday business. Two types of data could be generated here.

Of course, objective, end-result data. We're very interested to see if an error with that particular drug makes it out of the verification process.

But secondly, we're also interested to see whether the drug creates those process errors that we talked about. The way that we do that is that pharmacists and technicians carry what we call a self-monitoring booklet around with them, and whenever they catch themselves about to make an error with this targeted drug, we simply ask them, when they have a moment, to pull their booklet out and simply note a tally mark, oops, almost messed that one up. We also ask them to monitor those self-corrections for other drugs because we want to look for confusability pairs and see if any of those are there.

So both types, subjective and objective data, are recordable.

The advantage to the monitoring station is that there's really no conflict of interest in the sense that it's kind of a live test. We're not expecting any kind of result. We know maybe what we should see based on the earlier stages of the model, but there's really no hidden agenda ideally based in that. It's an actual, real-world environment, as realistic as we can make the simulation. That's the goal of the monitoring station.

There are marketing ramifications as well. Drug companies could get some information about how these drugs may be marketed in a different way than they currently are or would be. There could be some information that comes out of the simulation with respect to that.

The disadvantages. There is a risk of accidental dispensation, the risk being that there's an actual order for drug A, the test drug gets dispensed to that person by mistake. That risk is there. It could be correctable with observers on site from the testing authorities.

There is a use of self-report data, and the process errors are completely self-report. We know from just human nature that sometimes we are not very quick to recognize the fact that we almost made a mistake,

especially if that mistake is one that could have caused potential harm. So we have to take the self-report data with somewhat of a grain of salt.

And there is a lack of sample size possible because the number of these monitoring stations is probably going to be fairly small because of just the expense and the coordination necessary to create this kind of system. So can we really say that what happened in six pharmacies is going to happen in 60,000? That's an issue that we'll have to deal with.

Now, let me say a brief word about validation overall because I think the model in its entirety can be talked about very quickly and very simply with respect to validation. The nice thing about the model -- and it's a model that human factor psychology has used for years in determining the usability of products and human and computer interactions and those kinds of things -- is it tends to verify itself predictively. In the initial stages of the model, we develop predictive expectations on what we should see in the later stages. If we don't see that, we can then go back and refine or revise those predictions, collect more data. So the predictive validation process is kind of inherent in the model.

As far as construct validity, the question we have to ask -- and it's a question I've wanted to ask this

entire morning -- is what exactly are we looking for here. I think what our model is designed to target, as far as a construct, is error proneness. What we're looking at is how prone or how vulnerable is that particular name to confusion as an average statistic? When we define error proneness as the construct that we're targeting, then the model begins to make more sense because every step of the model then can be targeted toward answering the question, is this a mistake-prone name or is this not a mistake-prone name? That I think is a broader question. It goes beyond just the mere issues of similarity orthographically and phonetically, even though that is a component, but it's a broader question that may give us a more complete answer.

DR. GROSS: Thank you very much, Dr. Schell.

The next speaker is Dr. Sean Hennessy, Assistant Professor, Department of Epidemiology and Pharmacology in the Center for Clinical Epidemiology and Biostatistics at the School of Medicine, the University of Pennsylvania. Dr. Hennessy will talk about quantitative evaluation of drug name safety using mock pharmacy practice.

DR. HENNESSY: Good morning and thank you.

First, by way of disclosure of conflict of interest, I want to point out that I recently accepted an invitation to serve as an unpaid member of the Board of

Directors of Med Errors.

So I'm going to be talking about quantitative evaluation of drug name safety using close-to-reality pharmacy practice settings. A lot of what I'm going to be presenting is similar to what we just heard from Kraig Schell with the notable exception that I'm unburdened by any practical experience in the area.

(Laughter.)

DR. HENNESSY: So I'm going to focus more on the context in which information from such simulations can be done. In Kraig's diagram, this would probably line up with the movie set.

So first I'm going to talk about a big-picture view of drug name safety. How do we improve the process by making it quantitative or why might making it quantitative improve it? I'll briefly go over a model for measuring the error-proneness of particular drug names in a mock pharmacy setting and then talk about a research agenda.

So an overly simplified view of drug naming as it currently takes place is that there's a name. It goes through some evaluation process, as we heard earlier this morning. It's largely a qualitative evaluation process, and then there's some outcome. Either we accept it or we reject it. This is much the same process as you could use either for tomato soup or for Andy Warhol's art.

So the question is will we derive any benefit from making what is a qualitative process and depicted here as a black box, not coincidentally since many of the processes are not particularly well described, so they have that black box feature to them. So is there any benefit to making a qualitative black box process more transparent and more quantitative? So let me talk about the possibilities there.

So what might some potential benefits be of injecting a quantitative aspect to this? First is that we make the process more explicit and systematic. We use a fuller range of available information. We have transparency of data and assumptions. We acknowledge places that we're uncertain, and we identify knowledge gaps that then serve as areas of future research.

So then we need to ask the question, once we have the evaluation process, do we have enough information to make an accept-or-reject decision? What underlies this binary decision, go/no go, or is there really a spectrum of drug safety or error-proneness? And there needs to be some decision as to where the threshold is set on that spectrum.

So maybe it's really a rating that we need to have as an intermediate step between the evaluation process and the outcome. Certainly the rating in the middle, which is probably what I'll spend the majority of my time on,

should incorporate the probability of error. However, we need to ask is this enough. Are all medication errors created equal? There are some data that 99 percent of medication errors don't result in an observable adverse drug event. So should we focus on all equally, or should we focus on those that are more likely than others to result in an adverse drug event?

For example, is substituting erythromycin for clarithromycin, two antibiotics with similar spectrums, equally bad as confusing chlorambucil which is a chemotherapeutic agent with chloramphenicol which is an antibiotic?

So the rating may also take into account the consequences of the error in addition to the probability of the error. So under consequences of the error, that probably has multiple components too, the first of which -- and I'm echoing some things that were said earlier this morning, but not because I knew that they were going to be said -- one of which is the probability of error of an adverse event given that an error took place. And what are some factors that might go into that?

The first includes adverse outcomes from not getting the drug that was intended to have been dispensed, and we can get information from that presumably from the placebo-controlled trials that have been done demonstrating

the efficacy of the drug.

The probability of adverse events also depends on the identity of the drug that is mistakenly substituted which may be measurable empirically as I'll talk about in a little while.

And the third factor is the frequency of adverse events in recipients of people receiving the substituted drug. So given the substituted drug, what's the safety profile of that? And that should be known from pharmacoepidemiologic data about those drugs.

So in this rating, we have two factors, the second of which has two subfactors. So there's the probability of the adverse event, and then there's also the disutility of the adverse event under consequences of the error.

Let me talk about disutility for a minute. Disutility is defined as the value of avoiding a particular health state which is usually expressed on a scale between 0 and 1. This could be measured empirically by asking patients standardized questions. An example of this is presented here. This is disutility for outcomes of occult bacteremia going from everything to a very small disutility for just having your blood drawn to a very high disutility for death. I'd like to point out here that there are apparently things worse than death.

(Laughter.)

DR. HENNESSY: So one possible quantitative rating would be the probability of error times the consequences of the error, the consequences of the error being the probability of an adverse event given that an error occurred, multiplied by the disutility of the adverse event.

So then we have two axes. On the y axis, we have the consequences of an error. On the x axis, we have the probability of an error. You multiply those two things together, you get a severity rating going from blue, not so bad, to red, terrible. So you can get a bad severity rating either if you have a very serious event that occurs infrequently or a frequent event that's not so serious.

And here's Einstein discovering that time is actually money.

All right. So then in a process we need to ask the question, what settings do we perform this evaluation in? We could think about doing it in any number of settings: inpatient pharmacies, outpatient pharmacies, physicians' offices, nursing home settings. This list can go on and on.

So let me talk briefly about a model for measurement of some of these parameters in a mock pharmacy practice setting. So here's a photograph of a mock

pharmacy. These typically exist in schools of pharmacy, although they can be built for specific purposes as well.

What we can hope to gain from looking at a model like this would be both an empiric measurement of the probability of error, as well as get insight into what the consequences of the adverse event would be from knowing which drugs are mistakenly dispensed for the intended drug.

So some of the features of the close-to-reality simulated pharmacy practice include that it could be done in new or existing simulated pharmacies.

It could be done either using per diem real pharmacists or late-year pharmacy students, with the tradeoff being it costs more money to pay real pharmacists than it does pharmacy students, but you might get more realism.

The test drugs that we're studying would need to be listed both in the computerized drug information sources that are being used in the pharmacy, as well as in the computer system in which they're entering.

Then, of course, test drugs need to be put on the pharmacy shelf.

We would then simulate pharmacy practice by presenting prescriptions, phone prescriptions, electronic prescriptions, written prescriptions, for both the real drug and the test drug. As was mentioned earlier, you can

add prescription volume, noise, interruptions, third party reimbursement issues, Muzak, irate patients, as you like. The pharmacist enters the prescription into the computer system and then fills it. Then we measure the rate of name mixups at all stages of the filling process, as well as which drug was mistakenly substituted.

So when using the data obtained from such simulations to our formal quantitative evaluation process, we need to ask for the probability of an error. Do we use the measured probability of the error or do we use something else like maybe the upper bound of the 95 percent confidence interval? To remind you, the upper bound of the 95 percent confidence limit is the maximum value that is statistically compatible with the data and it's a function of both the study size and the measured rate, the point being that if we require use of the upper bound of the confidence limit, that will encourage a larger study than using the point estimate.

Which confidence intervals do we want to use? That might be subject to debate. 95 percent confidence intervals are common for biomedical research. It's a different context here, so we might want to think about other confidence limits, and that may be based on what seems reasonable going through this whole process with drugs that are at least assumed to be bad, some gold

standard bad drugs, if there is such a thing.

Potential advantages versus expert opinion. First, it yields empiric estimates of the error rate and of which drugs are mistakenly substituted. I would put forth it has better face validity. Further, the validity can be tested by examining known bad drug names, if we can get a group of people in a room to agree to what those are. It makes the knowledge and assumptions that go into the process explicit and transparent.

Obstacles and limitations. There are certainly those. The first is the Hawthorne effect; that is, when you watch people do something, they're generally better at it than when you're not watching them. The way to overcome that is if you do it enough, the Hawthorne effect is thought to go away.

There are technical challenges in developing movie set pharmacies and making them work also.

You need large sample sizes. Presumably these are going to be low frequency events, and in order to detect low frequency events, you need lots of repetitions. That's going to be expensive.

Do we use such a process routinely for all new drugs, or maybe do we use this as a way to validate existing or improved or otherwise less costly processes? And is doing so worth the added cost?

So now let me put forth the research agenda with regard to this particular proposed model. First is feasibility. Second, cost. Reliability. If we implement this strategy in different settings, do we get the same answers? The validity of it vis-a-vis what we believe to be both known good names and known bad names. And the ultimately utility of it.

So this is the straw man that I'm putting up for discussion, and I'd be happy to take any questions. Thank you.

DR. GROSS: Thank you, Dr. Hennessy.

At this point we'll entertain questions for Dr. Schell and Dr. Hennessy and Dr. Hennessy's straw man.

(Laughter.)

DR. GROSS: Yes, Jackie.

DR. GARDNER: I'd just like to ask Dr. Hennessy whether you have a recommendation about how routinely these should be used, given what you've described as a fairly extensive and expensive prospect. And if you only focused on the 1 percent of AEs that resulted in harm, for example, or targeted those, then you're looking at a big effort here. Do you have some modeling recommendation for how to decide what would be the most useful or cost effective way to proceed with this?

I was thinking of your Hawthorne effect not

only in observation, but proceeding with an IRB which would be necessary for this. Having described to everyone what exactly you're doing as part of the IRB process, then you'd have to wait even longer, I would think, before you saw --

DR. HENNESSY: Right. It's a cumbersome process. Is it worth it for all drugs? That's a good question. It's really a policy decision that I'll leave to the group for discussion.

DR. LEVIN: This is just a point of information. If there are no human subjects involved, why is this an IRB issue?

DR. GARDNER: Probably because the pharmacists' activities would be looked at. That would probably be the stance taken.

DR. GROSS: Michael Cohen.

DR. COHEN: If you're doing it in a live pharmacy, which at least one of the speakers talked about, there's always a chance of an actual error, and that has actually happened. We've had a recent report of a test of a computer system that led to a very serious error.

Could I ask a couple questions?

DR. GROSS: Go right ahead.

DR. COHEN: Has anybody actually used this model at this point, and is there anything in the literature about it? Because I think I'd like to know more

about it. I see some possibilities, but I haven't actually seen that used. Has anybody actually done this with proposed names, not with actual products that are on the market? That's the point.

DR. SCHELL: When you say this model, the whole entire thing or --

DR. COHEN: The model pharmacy concept. The lab is one thing, but the model pharmacy --

DR. SCHELL: Right. Not that I'm aware of. I'm currently speaking with a school of pharmacy right now about negotiating with them to use a new simulation that they're building, but to my knowledge, I don't know that anyone has done that.

DR. COHEN: I have one more question. When you do this, you would use actual handwritten prescriptions, but in fact, you'd need to test several handwritten prescriptions from different people that actually wrote that in order to make this work. So not only do you have perhaps 10 different drugs, but you might have 10 different actual scripts. It gets to the point where is this really a real-world experiment. That's the one concern I would have if you actually used a model pharmacy.

DR. SCHELL: And there's no question that as the model gets down toward the base of the pyramid, the complexity of it obviously dramatically increases. In an

ideal world, what we would hope is that the initial stages of the model would give us some idea about what sorts of script you might be more or less likely to see.

The other thing that I would say to that too is that, as you know, more and more scripts are now coming into the pharmacies electronically or with typewritten words, and also there's the whole bar coding phenomenon that's coming up. So I think that the model pharmacy will get less complex when that becomes more of a frequent occurrence.

DR. GROSS: Just to clarify, of the four simulations described, lab simulations have been tested, pharmacy schools simulations have been tested, movie set simulations have not, and real pharmacy simulations have been done. Is that correct?

DR. SCHELL: Let me say this to that. With respect to our particular research and research like ours, the laboratory simulation has been done and the field work which would be most similar to the error monitoring stations, at least a version of those -- we've done those in the past. But this particular model that I presented to you today in the context of drug name confusion is a synthesis of several different approaches that at this point is a framework model at best.

DR. GROSS: Any other questions? Yes, Ruth

Day.

DR. DAY: I'd just like to, first of all, express regret at the passing of Tony Grasha. He had so many creative ideas, and I'm pleased that Dr. Schell is able to continue his collaboration nonetheless.

My question for him is, as you go from the controlled laboratory situation to the real world, you're increasing ecological validity and decreasing control, but are there some controls that you can keep? For example, when a pharmacist has to go and find a particular drug that's a target drug, how many foils, that is to say, other things on the shelves, would there be? Is that the type of thing you can continue to control?

DR. SCHELL: Certainly. And in fact, you could even create that as a manipulable variable. What I'm reminded of is an experience we had with a chain in Florida who had created a targeted drug shelf, so the top 25 drugs that usually got misfilled, according to their records, were put on a special shelf with special markings and designated as different from other types of drugs that could have been confused as similar to it. Now, that particular intervention was not tested. It was just an idea somebody had and they decided let's just do this in the pharmacies. They really didn't have any idea as to whether it worked well or not. So, yes, that's one way

that comes to mind immediately when you say that, that you could test different kinds of targeting mechanisms, adjust foils, et cetera.

DR. GROSS: Eric Holmboe. Dr. Furberg.

DR. FURBERG: I also worry a little bit about comparability when you compare experimental settings to real life. I'm particularly concerned about whether the individuals you're examining know that they're being tested. They always do better. We know that from other settings that if you know you're being observed, you spend more time, are more careful, and you end up with an under-estimation of the problem.

DR. SCHELL: I think that's a valid concern and I think where that would be best addressed would be in the error monitoring stations with some sort of blind or double-blind procedure. That makes it a bit more complex to install and makes perhaps controlling the possibility of an error escaping the pharmacy more difficult to deal with. But that would be the solution to the problem.

Now, at the more basic levels of the model, I must make this very clear. My approach to these issues is slightly different than Tony's was. Tony's was very applied, you know, let's do the interventions and put them together right now, let's get them in the pharmacy. The reason he and I complemented each other so well is that I

tend to be more on the basic side. I tend to be more on the basic cognitive and perceptual factors that contribute to confusability in a broad context that then can be applied to the study of errors. So we worked very well together that way.

That's the part of the model that I think -- they're going to know they're being tested, and I'm not sure there's that much you can do about it.

DR. GROSS: Eric, did you have a question?

DR. HOLMBOE: No, I'm fine.

DR. GROSS: Louis.

DR. MORRIS: I had a couple questions for Dr. Hennessy. The idea of moving from qualitative to quantitative is very appealing, but in theory doesn't every drug potentially have a consequence and a probability with every other drug? So how do you go across when there may be so many drugs, and have you given any thought to how you might get the indices that represent the potential across the whole range of drugs?

DR. HENNESSY: So one way to do it would be you only take the drug switches that you observe empirically. They're the ones that you do the calculations for and assume are going to be the basis of your adverse event. So if you don't observe it, you assume it doesn't happen, which means that you need to do large enough studies.

DR. GROSS: Arthur Levin.

DR. LEVIN: I guess this is a question for both speakers. How do you design the simulation? There's a lot of range in choice in what the variables are and how you weight those variables. You know, do you have more Muzak and less angry customers? Is there any empirical base for sort of trying to emulate what the average setting might be, number one?

Number two, if there isn't, is that sort of a gap in data collection? In other words, if we're only getting reports this happened and there's very little detail, should we be looking for much more detail about the setting and the circumstance? I suppose that's part of the RCA maybe. But it seems to me if you build a simulation that purports to represent the real world, you better have some real-world foundations for putting that together.

DR. HENNESSY: I think that's a good point. I would probably do some observations in real life, quantitate those factors in real life, and maybe set the pharmacy at the 75th percentile of that, just as an example.

DR. GROSS: Michael Cohen.

DR. COHEN: Yes, that's close to what I was just going to ask. But I need to point out that the pharmacy is only one area that these errors actually occur,

obviously. A lot of it is on the nursing unit, in the ICU and the emergency room and the OR, et cetera. There are different environments. There are different types of patients. There are different jargons, et cetera. That would have to be taken into account because some of the worst errors we actually experience are in those very areas.

DR. SCHELL: If I could, let me speak to what both of our expert panelists have said and kind of piggy-back on what Sean said. Obviously, no simulator is perfect. Even the aircraft simulators they have in the Navy and the Air Force aren't perfect. They're awfully good, but they're not perfect.

Ideally -- and again speaking in either world -- the simulation in the later stages of my model would be built from data collected in the early stages of the model. I know that, for instance, there's currently being work done on things such as Muzak and other environmental factors by a company in Canada that I'm working with right now and the researchers up there who are doing good work right now in figuring out what environmental conditions impinge on performance and those kinds of things.

In the movie set and in the college of pharmacy portions of the model, as Dr. Day said, we can manipulate some of those things. For instance, when does music become

noise is a question that has to be asked. We know something about that factor from human factors literature, but we have not applied that basic knowledge to the pharmacy setting. We would need to do that to build the simulator effectively.

DR. GROSS: When music becomes noise is also relative to the listener.

Eric Holmboe.

DR. HOLMBOE: I have a question for both of you. There's been a lot of work also done in evaluating physician competence using simulation, particularly standardized patients. But at the same time, there's a growing body of work in actually videotaping encounters. And I'm thinking of the same thing with regard to pharmacies and other things. Has any work been done in that area where they've actually had ongoing video camera type analysis and break it down more, kind of an ethnographic type of study in those environments?

DR. SCHELL: I can only speak to the one piece of work that I'm familiar with. I'm familiar with it because we used it to validate our original laboratory simulation where pharmacists were filmed from the beginning of a script to the final production, primarily used in time motion studies. Dr. Lin at the University of Cincinnati has done a lot of work with shaving time off scripts and

looking at motion effectiveness and those kinds of things. We used that work as a validation for our own process to figure out whether we were able to reproduce the time it took to fill a script and approximately the number of errors that were being produced in those studies as well. But predominantly, to my knowledge, those were used in efficiency studies for the most part.

DR. COHEN: Can I help to answer that too?

DR. SCHELL: Yes.

DR. COHEN: There is some excellent work by Flynn and Barker which was the direct observation using video. So it was very revealing.

DR. SCHELL: Yes. Good point. Thank you. I forgot about that.

DR. GROSS: Paul Seligman and then we're going to break for lunch.

DR. SELIGMAN: Has there been any effort to compare the ability to detect the error proneness of a product in laboratory or simulated environments or more real-world environments with some of the other techniques that we heard about this morning using computer-based orthographic and phonographic techniques or expert panels? Have either you all or others had the opportunity to conduct those kinds of comparisons?

DR. SCHELL: Not to my knowledge. Dr. Dorr may

know of something. Maybe Mike might know of something. But from my reading of the literature, it's basically you have the computer approach and then you have the non-computer approach, and the twain have not met yet.

Ideally that's one direction I definitely want to go in. In fact, one study that I'm going to do. as soon as we get the drug name confusion lab constructed at ASU, is construct similarity indices and then run those pairings and those drug names through my perceptual task on the computer to see what kind of correlations I get. Do I get the kinds of proportions of errors that I should expect based on similarity ratings, or am I seeing a lack of correlation there? I think that would be very informative.

DR. GROSS: Okay. Thank you all. It's been a very interesting morning. We will break now and we will reconvene at a quarter of 1:00, 12:45. Thank you all.

(Whereupon, at 11:40 p.m., the committee was recessed, to reconvene at 12:45 p.m., this same day.)

AFTERNOON SESSION

(12:45 p.m.)

DR. GROSS: We will begin the open public hearing. For the panel, you have a purple folder that has much of the information that will be presented. Patricia Staub will go first.

MS. STAUB: Good afternoon, ladies and gentlemen. It's a pleasure to be here today on behalf of Brand Institute to present to you --

MS. JAIN: Patricia, could we just hang on just one second. There has to be a statement that's read first. I apologize.

DR. GROSS: Before we begin, I have the pleasure of reading a nice, long paragraph to you.

(Laughter.)

DR. GROSS: Both the Food and Drug Administration and the public believe in a transparent process for information-gathering and decision-making. To ensure such transparency at the open public hearing session of this advisory committee meeting, FDA believes that it is important to understand the context of an individual's presentation. For this reason, the FDA encourages you, the open public hearing speakers, at the beginning of your written or oral statement to advise the committee of any financial relationship that you may have with any company

or any group that is likely to be impacted by the topic of this meeting.

For example, the financial information may include a company's or a group's payment of your travel, lodging, or other expenses in connection with your attendance at the meeting.

Likewise, FDA encourages you, at the beginning of our statement, to advise the committee if you do not have any such financial relationships.

If you choose not to address this issue of financial relationships at the beginning of your statement, it will not preclude you from speaking.

So the first speaker is Patricia Staub.

MS. STAUB: Good afternoon, ladies and gentlemen, once again. It is a pleasure to be here today on behalf of Brand Institute to present to you several key issues and recommendations with respect to minimizing the risk of confusion caused by look-alike and sound-alike proprietary names for branded prescription drug products.

By way of introduction, I am a licensed pharmacist and attorney and a former FDA employee. I am currently employed as Vice President of Regulatory Affairs for Brand Institute. Brand Institute is a well-known and experienced international brand development company that routinely conducts name confusion studies and makes risk

assessments in the process of developing proprietary names for prescription drug products.

During the past five years, Brand Institute has participated in the brand name development of nearly half of all the prescription drug brand names approved for use in the United States.

On behalf of both Jim Detorre of Brand Institute, the CEO, and myself, I thank you for inviting us here today to share with you our own best practices and recommendations relative to the brand name selection process. If there is time at the end of my talk, I'd also like to briefly address the five questions before the committee and give you our opinion on these five questions.

Recognition and memorability: benefits versus reality. The hallmark of a successful proprietary name is high brand recognition and memorability. Easily recognizable and memorable names may, indeed, sell more product, but strong brand names are also safer names, ones that are less likely to be inadvertently confused with other drugs. Therefore, we all struggle to provide safer brand names that benefit both prescriber and patient by decreasing the risk of medication errors associated with look-alike and sound-alike names. This is no small challenge today with over 17,000 brand and generic names approved in the United States alone, and only 26 letters in

the English alphabet. Given these statistics, some similarity between drug names cannot be avoided. Our objective then is to avoid confusing similarities between brand names.

When brand names are found to be likely to cause confusion, one way to manage the risk of medication errors is to increase a brand's recognition and memorability. Some of the newer methods may involve promotional campaigns around drug names after they're on the market.

Risk management techniques. Pre-approval methods of managing the risk of medication errors due to brand name confusion have surfaced in the relatively recent past. Regulators in the wake of the 1999 Institute of Medicine report, To Err is Human, have increasingly sought to shift the burden of risk management for brand name confusion to industry.

Today when a pharmaceutical company proposes a brand name for their soon-to-be-approved drug, the agency, through DMETS, will review that name for safety. The results of prescription interpretation studies which assess the risk of brand name confusion and the potential for patient harm have become part of industry's routine activities in bringing a brand name to market. Also during the pre-approval period, sponsors have started airing

"coming soon" ads to get the name out to the public, thereby increasing recognition and memorability of new drug names.

Proactive post-approval risk management activities can be particularly useful in that initial period immediately after a drug's approval when prescribers may be unaware of the new drug name and the risk of medication error can be high. Reminder ads as part of a strong launch and targeted advertising are also employed to increase name recognition. When name recognition fails and confusion occurs, Dear Doctor letters informing physicians of the confusion of names, the use of tall man letters to accentuate differences in product names already on the market can be helpful. Name withdrawal should be a last resort.

With these thoughts in mind, we would now like to share with the agency and the committee some of our own best branding practices developed through our experience and research at Brand Institute. We will then end with a few specific recommendations that we suggest to improve the regulatory review process for brand safety.

Best practices: multi-factorial real-world approach. While generating safety signals through a retrospective review of past errors can be helpful, we suggest that there is no substitute for using a multi-

factorial approach to generate potential safety signals associated with the introduction of a new proposed prescription drug name. We believe that real-world testing among a large sample size of currently practicing health care practitioners is critical in addition to testing through orthographic and phonetic analysis, expert focus group review, impact review, and computer-aided research. Very often in doing this extensive testing, we do uncover strong signals in one category or another that causes us to reject a brand name candidate before it is submitted to the FDA. Our premise that this combination approach offers the most comprehensive and reliable methodology for confusion testing among brand names appears to be supported by our relative lack of confusion over the past couple of years when you compare the names that we've generated to the USP drug list.

Although differences of opinion regarding the results can still exist between regulators and sponsors, even when extensive testing has been completed, the inherent value of this testing is that awareness of risk is identified and monitored. And risk management strategies may be employed by the sponsors and the agency either prior to marketing or as a condition of marketing their product under their preferred brand name. Once a potential risk is identified, it can be qualified and hopefully minimized.

Lessons learned from AERS. A retrospective analysis of all reported mortality-associated medication errors contained in the AERS database during a 5-year period ending in 2001 was published on the CDER website. Jerry Phillips' group was the author of this study which looked solely at fatal medication errors, the most serious consequences.

It is interesting to note that the confusion rates of brand names were similar to the confusion rates of generic names, that more written miscommunications rather than oral miscommunications resulted in fatal errors, that elderly patients over 60 years old in hospital settings receiving injectable drugs for CNS, oncology, and cardiovascular conditions were more frequent victims of fatal medication errors. Most patients that died were taking only one medication according to the study. These potentially predisposing factors should be considered possibly when assessing brand name risk: patients again over the age of 60 in hospital settings receiving injectable drugs and particularly patients taking therapeutic categories of CNS, oncology, and CV.

10 percent of these medication errors were fatal, of the 5,366 that were measured, and the most common error was an improper dose, 40.9 percent. The wrong drug was 16 percent of the time, and wrong route of

administration, 9.5 percent of the time. Proprietary name confusion resulted in 4.8 percent of the medication errors, and nonproprietary name confusion resulted in 4.1 percent of the fatal medication errors. 6.7 percent were due to written miscommunications and 1.7 percent of the fatal errors were due to oral miscommunications. 48.6 percent of the deaths occurred in patients over 60 years of age, and the largest number of deaths, 26.7 percent, occurred in the practice setting of a hospital. The most common dosage form again in death due to medication orders was injectables, 49.9 percent.

Benchmarking. Benchmarking is a topic where we have a lot of questions from our clients. We believe that benchmarking error rates in confusion studies, while relevant, can also be misleading without a separate evaluation of the impact on patient harm. For example, even high error percentages based on potential name confusion with another drug whose misadministration would likely result in little or no patient harm may not be as meaningful as a much smaller error rate percentage that would likely result in high patient harm, for instance, mistaking a diuretic for an oncology product.

Benchmarking, combined with impact analysis, is a more useful tool for assessing risk.

Another misleading aspect of over-reliance on

benchmarking can be the fact that a certain number of errors in confusion testing may be the result of misspelling the new name rather than confusing the new name with another drug. Misspellings alone may be harmless.

Overlapping characteristics. Brand name similarity cannot likely be completely eliminated due to the large number of approved brand names in the United States. Similar or overlapping characteristics, however, in combination with a similar brand name, can be important additional causes of confusion, and these characteristics should also be evaluated in brand name confusion studies. For example, similar packaging, labeling, route of administration, dosage form, concentration, strength, patient settings, storage conditions, and frequency of dose may make a difference between a similar brand name and a confusingly similar brand name. In our brand confusion studies, we prepare a chart that looks at overlapping characteristics between similar sounding and looking names as a factor in making our risk assessment for name confusion.

I guess they're going to exclude modifiers from this setting. So all I will say about that is that with the general policy that the agency has that only one brand name per product per sponsor will be approved, brand name modifiers are the only way that a manufacturer can use to

further define new formulations of their product. Of course, there are problems with modifiers that are well known. In Europe prefix modifiers are sometimes used and because of our international business, sometimes clients would like to have prefix modifiers. This can really create problems I think in the United States, and I'm glad that we don't have a problem with people suggesting prefix modifiers here.

The suffix modifiers everyone knows are problems due to the fact that XL and SR have a variety of meanings, depending on the drug product that you have. In Europe if a drug modifier or suffix modifier doesn't have the same meaning in each of the member countries, it's not allowed.

Particularly the suffix XL I think, should be noted, can be confusing with the quantity of 40 tablets, since that's the Roman numeral. There are several two-letter suffixes that are problematic. One-letter suffixes are not allowed in not allowed in Europe, and I think that they're fairly rare in the United States too. That's probably a good thing because modifier drop-off is probably more prone with the one-letter modifier.

On the subject of numerical branding, numerical branding is using numbers in a single entity brand name, and we highly discourage this in general since the name can

be confused with the strength or dosage. For instance, valium-5 can look like take 5 tablets of valium and can result in medication overdose. Numerical branding for combination products, however, can minimize confusion and improve safety in some cases but only if both ingredients are listed numerically. For example, referring to Percocet 5, oxycodone 5 milligrams/acetaminophen 325 milligrams, by only it's oxycodone number 5 can lead to the administration of 5 tablets of Percocet and cause fatal patient harm. However, referring to Percocet without the number 5 or only using the number 5 in conjunction with the number 5/325 can make clearer the dose required.

Trailing zeros. We agree with ISMP that trailing zeros can cause confusion and that brand names should never be accompanied by dosages with trailing zeros. For instance, 2.50 milligrams can be interpreted as 250 milligrams. Leading zeros, however, do improve the absence of confusion and should be always used. 0.25 milligram versus .25 milligram.

Tall man letters. The use of capital letters within a generic name to differentiate nonproprietary names, acetaHEXazole and acetaZOLamide, is one risk management technique that could be applied to brand names in the post-marketing setting to differentiate them. This has been done recently with SeroQUEL versus SaraFEM and

SerZONE. And that's an example of SeroQUEL's new packaging that accentuates the difference between confusingly similar names.

Bar coding, while we recognize its importance, only has limited importance. It minimizes order picking confusion, but does not minimize interpretive confusion. Computerized order entry may minimize illegible handwriting from prescribers, but it also may introduce its own set of errors in picking a drug from the list. Electronic solutions to these problems are not totally error-free.

Orthographic analysis, looking at strings of letters, are instructive, but this method alone does not adequately address confusion. Orthographic analysis may be more helpful in real-world, handwritten prescriptions as it can show the formation of certain letters may decline in somewhat predictable ways such as an M bleeding into an N.

We also agree with DMETS that beginning drug names with the letter Z can be problematic in that Z, when scripted, may look like C, L, B, 2, g, y, j, or q, and might sound like C, S, and X.

We have three recommendations for the process of naming that we would like to make.

The first suggestion that we have -- and this is really a result of some of the problems that we've experienced with our clients during the process -- is that

tentatively approved names be made public, when they are tentatively approved, via the internet so that successive name candidates can test their own proposed proprietary names against names that have already been tentatively approved, but could potentially beat them to the market. Confusion testing is only as good as the universe of names that the proposed name can be tested against.

The second suggestion we have is that whatever testing models DMETS uses from time to time, that those testing methods be made transparent so that comparison between the two models can be made and parallel testing of names could possibly improve the accuracy of both models, both the proprietary model that was being submitted to the FDA and the FDA's own model that it's testing.

A third issue that we would like to suggest is duplicate brand name exception for drugs where the brand name is already widely associated with the treatment of mental illness and stigma has been proven and a second drug name possibly should be allowed for that compound where there is a physical illness. Wellbutrin versus Zyban and Prozac versus Serafem are two examples of this type of exception to the normal rule of one brand name per drug per sponsor. We believe that if stigma can be proven, patient harm can be alleviated that may be caused by embarrassment for taking a well-known mental health drug for a physical

condition, particularly where employer-paid prescriptions are available.

In conclusion, there are many opportunities during the name development process to safeguard against medication errors caused by look-alike and sound-alike proprietary names. High recognition and memorability are key components of safe drug names. While post-marketing risk management programs are useful, pre-marketing activities are increasingly being used to anticipate and identify risks before harm occurs.

Although predicting risk is not an exact science, neither is medicine. Human error is a predictable constant in any health care system. No medication error prevention technology is itself error-free. A multi-factorial, real-world approach to names testing to prospectively identify levels of risk associated with new drug names during the approval process is key.

We applaud the efforts of the agency in taking up this difficult challenge to patient safety by creating the DMETS layer of brand name review and attempting to establish patterns by retrospective analysis of the AERS database. While differences of opinion may still exist between regulators and sponsors as to levels of acceptable risks associated with a drug name, we do not see any realistic substitute for comprehensive name testing in the

real world to assess the risk of confusion between new and existing drug names. After all, the prediction of risk is always based on probability and is never absolute. Real-world testing allows us to observe risks that have already been seen rather than to speculate on risks that may occur.

Thank you.

DR. GROSS: Thank you.

There are four more presenters. We would like to finish these remarks by 2 o'clock. So I would ask the other presenters if they could condense their presentation a little bit.

The next speaker is Dr. Douglas Bierer from Consumer Healthcare Products Association. He's Vice President of Regulatory and Scientific Affairs. Thank you.

DR. BIERER: Thank you. Good afternoon and thank you for the opportunity to present an OTC perspective on sound-alike/look-alike drug names. While OTC products are not the subject of this panel's conversation today, it would be important to mention some comments about OTC drugs since they were mentioned briefly in this morning's presentations.

The Consumer Healthcare Products Association, which was founded in 1881, is a national trade association that represents the manufacturers and distributors of over-the-counter drug products, and our members account for more

than 90 percent of the OTC products that are sold at retail in the U.S. CHPA has a long working history with the FDA to improve OTC labeling so that these labels are easier for the consumer to both read and understand.

In considering the issue of drug names for OTC products, it is important to stress several key differences that arise from both prescription and OTC drugs. One of the most important differences is how the drugs are purchased. Prescription drugs are made available by written or verbal order by a physician or a licensed practitioner, which then, in turn, needs to be translated and filled by a pharmacist.

OTC drugs, on the other hand, are purchased directly by the consumer. Thus the OTC product package must communicate all of the information the consumer needs to decide if it is the right product for them. When purchasing an OTC medicine, the first thing the consumer sees on the store shelf is the product's principal display panel.

As shown in this slide, in addition to the brand name, the principal display panel includes other important information to help consumers identify if it is the appropriate product for the condition that they want to treat.

First is a statement of identity. This

includes the established name, that is, the official name of the drug and the general pharmacological category or its intended action of the drug. It is written in layman's language and must be prominent and conspicuous on the package. And for those products which are combinations of active ingredients, there must be a statement about the principal intended action of each of the active ingredients. All these elements are required on OTC packages.

Often the principal display panel contains other information such as the dose of the active ingredient and perhaps a statement about a product's benefits, such as it relieves or treats a certain type of ailment.

In addition, it may contain a flag in the upper corner to alert consumers of important new information. This flag was a voluntary program first initiated by CHPA in 1977 to provide consumers with more information when they were purchasing OTC drug products. In this case the flag says "new," indicating that this is a new product. It may also say "see new labeling" or "see new warning" to indicate that a change has been made to the product labeling on the back of the package.

All of this information is clearly visible at the point of purchase and helps the consumer to decide if this is the right product for them.

The next major difference is the drug facts labeling. By May 2005, all OTC medications will be required to use this format, and in fact, many OTC products are already using them on the store shelves. Drug facts standardize all the labeling on the back of the package to make it easier for the consumer to read and follow the label. The information appears in very clear, concise consumer language. As shown on this example of a chlorpheniramine product, the drugs facts includes the active ingredient of the product, including the quantity of each active ingredient per unit dose, the purpose of the active ingredient, what the product is to be used for, any warnings about the use of the product which are grouped in headers to facilitate the consumer finding the information and understanding the information.

Next, the directions, which is important to mention that the directions appear after the warning signs in an OTC package.

Finally, other information such as storage conditions, and finally a list of inactive ingredients listed in alphabetical order so the consumer can know what is in the product that they're going to be taking.

Because this information is organized in exactly the same way on every OTC product, this format makes it easier for the consumer to find all the

information they need to take the product correctly and safely and also when to contact a physician. It is also important to note there is redundancy of the information in drug facts and on the front panel, and this serves to reinforce the information sent to the consumer.

At the 26th June meeting on drug naming and also at this meeting, the agency expressed concern about OTC brand name extensions in which a family of products may have a similar name and may be used for different conditions and may contain different active ingredients. OTC brand names allow consumers to locate a family of products which they have used before and that they trust. OTC manufacturers confine the family of products to particular therapeutic areas in order to decrease the concern that consumers may take a product for one condition when it really should be used for another condition.

It has also been suggested that brand trade name extensions should not be used and that each extension should have a differently named product. However, this approach has potential to create more consumer confusion because the consumer will be required to master separate information and brand names for each product. As these products are advertised in the media, the plethora of different products will create confusion and make it even more difficult for consumers to remember what the product

is to be used for and for what conditions. Brand names and their line extensions do provide consumers with valuable information about the products that they have used before and that they have come to trust.

As I have illustrated, the consumer has much more information than just the brand name to recognize when selecting an OTC product. The uniqueness of the amount and the redundancy of the information on the OTC label, when compared to handwritten or oral prescriptions and prescription product packages themselves, decreases the reliance on the brand name and aids the consumer in making the right choice about the product for the condition that they want to treat.

Thank you for considering the views of the OTC drug industry.

DR. GROSS: Thank you very much.

The next speaker is Clement Galluccio from rxmarx, a division of Interbrand Wood Healthcare.

MR. GALLUCCIO: No slides for me today. Just I guess the burden of having been involved in the validation of proposed pharmaceutical trademarks for close to 15 years. I guess that's in opposition of being unburdened of no practical experience.

In 1991, Interbrand Wood Healthcare and rxmarx introduced the 10/10 trademark evaluation model to

immediate acceptance from many of the world's leading pharmaceutical companies. Of the many innovations introduced with the 10/10 model, paramount was the concept that trademark selection was more complex than the exclusive consideration of prescriber preference, but also reflected the desire to select a safe name. To date, over 80 trademarks have been first 10/10 certified prior to agency submission and subsequently introduced to the marketplace, with many more presently waiting introduction.

To the best of our knowledge, less than 2 percent of trademarks validated using the 10/10 model have encountered any degree of concern relative to medication error. These 80 trademarks are representative of over 700 name validation studies, consisting of thousands of proposed pharmaceutical trademarks.

Given the significant role that Interbrand Wood Healthcare and rxmarx have served in creating and validating pharmaceutical trademarks, there have been many important lessons that we have learned in regard to the identification of names at risk of medication error. The one that we most often share with our clients in regard to the certainty of our findings is the following. Regardless of the methodology used to validate a pharmaceutical trademark, each and every name has the potential to be communicated so poorly by the prescriber or transcriber

that it could be potentially mistaken for another product name.

Therefore, it stands to reason that unless significant changes are made to how pharmaceutical products are packaged, distributed, stored, and communicated within the dispensing environment, independent of changes to validate nomenclature, medication error will continue to be a harsh reality for all concerned. Minimizing medication error, not finding alternate methodologies to validate proposed pharmaceutical trademarks, should be the primary focus of the discussion. That said, it is the opinion and recommendation of Interbrand Wood Healthcare and rxmarx that both industry and agency should strongly consider the following.

Grant equal time and consideration to the factors other than trademark similarity that may also contribute to medication error. As David Wood, CEO of Interbrand Wood Healthcare, shared on June 26th, let's not make trademarks the whipping boy for a system which needs to pay attention to the many other things other than the brand name. A good start would be to begin validating nonproprietary names for safety using the same best practices that have been developed for proprietary names, followed by paying much closer attention to labeling, packaging, and administration practices. Perhaps the

answer to minimizing medication error exists in creating greater personalization, differentiation and security in product labeling, packaging, and delivery systems as opposed to creating increasingly more restrictive barriers to proposed pharmaceutical trademarks.

Two, fund a study to provide an accounting of previously identified nomenclature associated with medication error over the past 10 years, as well as determine present nomenclature assessment practices by sponsors. We believe there exists a significant absence of data relative to the actual as opposed to the perceived causes of medication error. The anticipated outcome would be to better understand which factors, for example, brand name versus generic name, the lack of adequate legal or research assessment prior to introduction, overlap of dispensing profile and other dispensing factors and practices, et cetera, that may have significantly influenced medication error.

Three, in recognition of the many companies within industry that have already implemented best practices relative to nomenclature validation, provide flexibility within whatever guidance, whatever outcome to follow to allow such companies to continue in their present approach until new methodologies are validated. In our view the best practices for the validation of proposed

pharmaceutical nomenclature already exist, however, need to be applied on a consistent basis by each and every sponsor. In turn, agency should provide a predefined set of consistent metrics relative to approval or rejection so that the outcome of nomenclature validation studies is predictable, for example, a proposed name misinterpreted more than once for the same potential conflict is automatically determined to be of high risk or higher risk. High-risk candidates would then be considered for more in-depth analysis, perhaps quantitative analysis or monitoring programs post-launch.

In conclusion, we believe an inclusive approach is paramount in order to provide the desired benefit to the public in regard to minimizing medication error. We applaud today's participants for their efforts and agree that the development and selection of a pharmaceutical trademark should reflect best practices relative to the identification of a safe trademark. However, recent advances such as the increasing use of computer-assisted prescribing and dispensing tools is only one initiative that supports a more comprehensive approach. These advances, when combined with many of the existing best practices relative to nomenclature validation, as reflected in present methodologies and the recommendations I shared earlier, represent the most logical resolution to

minimizing medication error.

In conclusion, beyond our statement, we have released our methodologies, both proprietary and nonproprietary, to the committee so we can open-source these methodologies for use by all.

Thank you.

DR. GROSS: Thank you, Mr. Galluccio.

The next speaker is Maury Tepper III from Womble, Carlyle, Sandridge & Rice.

MR. TEPPER: Thank you and I'll start with what is a customary gesture for me: adjustment of the microphone.

I welcome the chance to be here with you today, and I do want to mention just a couple of quick things by way of introduction for you. I do share one thing in common with you members of the advisory committee. I am a special government employee as well for the Department of Commerce. I serve on the Trademark Public Advisory committee for the U.S. Patent and Trademark Office. My comments today will not relate to the Patent and Trademark Office or its operations, but I did want to make you aware of that.

I also, very importantly, want to note that I'm pleased to see that the ACC is well represented here. As a resident of North Carolina, I'm glad to see participation

from others who may also be traveling back to our State under a weather advisory today.

(Laughter.)

MR. TEPPER: I come to you I think bringing good news and hopefully some recommendations. And let me just step back as one who has previously served as in-house trademark counsel for a pharmaceutical company -- and currently I'm in private practice representing all types of clients, some in the pharmaceutical industry, some in industries such as snack foods, candies, and racing memorabilia -- and tell you that I think the good news here is everybody in this room shares a common interest and common goal. That is not always the case, but hopefully it has come through today. If it hasn't, I really want to emphasize I think both the FDA and sponsors are working very hard here, striving to do everything that can be done to find ways to minimize medication errors, to bring out the safest possible products, including their trademarks.

I think where we may differ is in determining how best to go about that and the degree to which trademark analysis contributes significantly to the problem or indeed may be the best solution. And I'll talk about that a bit in my remarks. But I think it's important to keep in mind and to understand here that at the end of the day, we're all seeking the very same thing. So I think the efforts in

this room are laudable. I think the fact that we're all working towards the same goal is encouraging and should mean that we can arrive at a very workable system or continue to refine that. I hope that this will be the lead-in to an open dialogue.

It is important I think to note in looking at this problem -- and I was very pleased to see some of the good questions this morning -- that a lot of the presentations, a lot of the data presented today start from an assumption that trademarks contribute substantially to medication errors. I think we would all agree that they are involved and that they are a factor, but I do have to reemphasize I'm not aware of any study or any way that we have come about determining what a significant factor they are or what their role is, if they cause the error. The fact that two name pairs are similar certainly doesn't automatically mean in every case that is a significant contributing cause to the error.

I was very taken by Dr. Dorr's research this morning in her presentation. For a dumb lawyer like me, it was the closest I've come to understanding some of that science, but it leapt out at me that in listing for you a degree of name pairs that had high similarity rankings, some of them were involved in errors, some of them weren't. That tells us that similarity alone is not the decisive

factor. It is not in all cases going to tell us automatically is this a problem. It is relevant. It is absolutely something we need to consider, but I would submit that it is simply one of many factors that need a balanced approach in making a determination about the safety of a name in its appropriate setting and context.

The other thing I think we need to be mindful of -- I liked Mr. Woods' characterization that was just quoted of not making trademarks the whipping boy for other parts of the system -- is to be thinking about where we can have the most significant impact on this problem.

You were shown this morning Avandia and Coumadin as two names that are somewhat similar. Of course, the only similarity there is in handwriting, and I do have to ask the question, if we were coming to the point where we're looking at trademarks as the part of the system to make up for sloppy handwriting, are we really getting at the problem in the best way? Are we going to have the maximum on patient safety by trying to do that? That's not to say we are not going to continue to strive to predict and identify and address these issues and create safe marks, but I think it is important to keep in mind that there are probably other more significant causes that we should be focused on and should be addressing as part of this effort beyond trademark review, simply because

trademarks are prominent and are identified in each situation.

The other thing I think is important to realize here -- and this is a scientific group. Again, as a dumb lawyer addressing you, I need to be careful, but at the end of the day, these are subjective determinations. We would love to have a validated test. We would love to have an objective measure that would tell us all whether or not we are going to have problems given a particular trademark. I have to tell you I simply do not believe that can happen. There are too many factors involved in each situation, in each setting, in each combination of drugs that come into play that need to be considered and need to be carefully weighed and need to be looked at to allow us to simply come up with a formula or any one approach that will give us some prediction of error propensity.

All of the techniques here that have been discussed this morning I think provide very useful data, but it's important to keep in mind that that's all they provide. They are sources of data. I don't think we have any one outcome predictor here. I applaud the efforts to continue to seek one, but I want to be careful here to indicate that we should best view these as inputs right now.

Another piece of good news for you I think is

to note -- and the question was asked -- you'll be getting some additional information about this, but just the degree to which trademarks are carefully screened and reviewed by both pharmaceutical companies and by the FDA. I can tell you as someone who works for clients in lots of industries, there is no industry that even comes close to the pharmaceutical industry in the care that it gives in the selection and consideration of trademarks. I get lots of calls from clients that are launching products next week. Thankfully, those tend to be snack cakes rather than drugs. Drug names are typically given very careful consideration. You'll hear more, and I think you heard from Bob Lee already this morning about the types of testing. But I think if you really break it down and look, the types of testing that FDA and that sponsors are engaging in really have a lot in common. In many ways they approximate one another.

Where I think there is a significant difference is in what is being done with that data. I would propose -- and my paper goes into this in some more detail that one thing we need is a framework for making decisions. All of these resources we've talked about this morning are best viewed as providing relevant data to you, but we need some framework for analyzing that data. The trademark system provides that.

I'll apologize for anyone who heard me on June 26th if I sound like a broken record. This is in some ways echoes my comments at that point. If anything, the outcomes of that meeting solidified that belief that given all of this data, that we cannot validate it and we need to decide what place it should have in each situation.

The best test is one that can carefully look at and approximate market conditions, and that is precisely what the legal test for trademark availability is designed to do. The likelihood of confusion test that is employed by attorneys, that is employed by the Patent and Trademark Office in reviewing proposed trademarks, that is employed by courts in determining disputes and whether there are actual conflicts is a test that is well established, well defined, and yes, it is subjective, but it is a well understood language for having this discussion and for analyzing and balancing these factors in each situation.

What makes pharmaceuticals special? Certainly this is a very different market than the consumer marketplace. In some ways it's frightening that the average consumer may go out and pay more attention and be more involved in selecting their laundry detergent than in receiving a medication where they in many ways turn it over to the providers and the dispensers and take whatever is handed to them in blind trust. We need to understand and

take those market conditions into account that the same test provides the ability to balance those factors, to use this input data about similarity orthographically and in handwriting and in sound, and to consider them in a framework that provides us something of a useful and predictable result that gives us the basis for analyzing these and for balancing the numerous factors that come into play rather than seeking to emphasize one single measure.

I do want to come back to the important notion, though, that as we are engaging in these efforts, I think that the FDA has done a laudable job in bringing focus to bear on the science available here and helping refine and establish some of these techniques and seeing how they're put to use. I think part of where we perhaps differ is once that data is generated, how is a decision arrived at. Attorneys are used to using a defined and documented and, I'll say, reproducible test to sort of have that discussion and make the analysis. FDA is looking at the same data and coming to conclusions. I think anytime you're dealing in a subjective area, that's natural and understandable. You heard Dr. Phillips I think this morning acknowledge sometimes when they have concerns, they turn out to be borne out in the marketplace, sometimes they don't.

Again, I wish we could give you an objective measure that's going to be a crystal ball for us, but I

think what we need to strive for is to make sure that good naming practices are followed, to make sure that these techniques have been employed and have been considered, and then recognizing that these are subjective judgments, to really carefully consider whether substituting the FDA's judgment for that of a sponsor is going to substantially increase or improve patient safety.

In many ways I submit that there are times when you may increase risk by causing a sponsor close to launch to have to go back and change a trademark. Typically trademark reviews -- and again, I'll echo Bob Lee's comments this morning -- occur at multiple stages. Certainly during the creation, the sponsors are generating these names and screening them internally. They're conducting an analysis. They're seeking input from appropriate experts. When the application is filed, the trademark is again reviewed by an examiner at the Patent and Trademark Office who is employing the same likelihood of confusion standard. Indeed, the Patent and Trademark Office and courts have both recognized a higher degree of care for pharmaceutical trademarks given the significance of similarity here.

Finally, the opposition period comes up and that's when competitors also conduct the same review, step in and oppose the mark if they feel there's a potential for

conflict or the mark is too close.

This process takes several years to complete, and so by the time a trademark application is filed or has been screened and filed, has been subject to an opposition, we have a lot of eyes over that that have come to some consensus that this mark does not appear likely to cause confusion. To step in and have to change that mark, without the time to go back through that process, in some ways deprives those others of the right to review and comment, forces the sponsor to make some last-minute changes or determinations, and to do their best, of course, in analyzing this. But I submit that we may be increasing risks in some ways by causing these changes close to launch and without the availability of these other reviews and mechanisms and considerations that we typically would want to employ.

I will leave my comments there in the interest of your time. I have provided some answers to the questions in the written material to you, but in large part, I think the key answer here is we need to continue to do everything we can to refine the techniques for generating information to consider, but we need to keep in mind that at the end of the day each of these tests can only provide relevant data that we should consider. This will be a subjective determination. There is a well-

established test that is used for making that subjective determination. Trademark attorneys have expertise in doing that and they attempt to balance the appropriate factors. I think FDA should continue to play a role in shaping practices that will provide the relevant data, should provide good naming practices, should ensure that industry is taking these into consideration.

I think FDA should be very cautious, however, at substituting its subjective judgment based on a standard that we do not know for that that has been arrived through the likelihood of confusion analysis.

I also think that we need to continue to do what we can to focus on the overall problem of errors, understand that trademarks are a factor, but also understand that efforts that may have greater impact and greater significance should certainly not be overlooked in the haste to squeeze tighter down on the most visible aspect of the system, and that is the trademark.

Thank you.

DR. GROSS: Thank you very much.

The next speaker is Dr. Suzanne Coffman, who is Product Manager of NDCHealth.

DR. COFFMAN: Thank you, Dr. Gross, members of the committee, and the FDA, for the opportunity to appear before you today. You should have a copy of my

presentation in your packet.

As Dr. Gross mentioned, my name is Suzanne Coffman. I am a pharmacist and I am a product manager for NDCHealth where my responsibilities include clinically based transaction products for the pharmacy market. In the interest of full disclosure, I'm also a shareholder of NDC and they did pay for my travel.

I spoke on this topic at the joint ISMP/PhRMA/FDA meeting in June. Today I'll be providing an update and also just expressing NDCHealth's continued interest in the topic of preventing drug name confusion errors.

NDCHealth is a leading provider of point-of-sale and information management services that add value and increase the efficiency of pharmacy, pharmaceutical manufacturing, hospital, and physician businesses. Two out of three prescription transactions in the United States travel across our network, and we are connected to 90 percent of retail pharmacy outlets in the U.S. We also process transactions in Canada.

One of the services that we offer to the retail pharmacy market is real-time alerts about drug name confusion errors. This service is supported by a database that contains all of the known look-alike/sound-alike pairs that involve oral solid products that are used in the

retail environment. To that list, we add a likelihood score, a clinical significance score, absolute dosing for each drug dosage form strength that is involved in the pair and also typical dosing for each form strength, and we derive that from the 160 million transactions that travel across our network each month.

We send an alert when the dose that is submitted on a prescription is atypical for the drug that is submitted, especially when it's typical for one of the look-alike/sound-alike pairs. This does reduce name confusion. Through our ability to match prescriptions and to look at the follow-up prescriptions, we have identified a number of changes, of course, in quantity and day supply, but we've also identified several changes to the drug. Some of these are known look-alike/sound-alike drugs; many are not. We've had changes, for example, between sartans and between ACE inhibitors which are not on the list.

We've also recently completed data collection on a randomized controlled trial in a regional chain, 115 stores. Preliminary results show that pharmacy staff, pharmacists' and technicians' knowledge of look-alike/sound-alike pairs did improve after exposure to our real-time alerts. However, even after exposure, they would have only made a C if they were taking a test in pharmacy school.

We are currently analyzing the data on the actual error prevention, again using our prescription matching methodology, so that we're able to tell what happened after the pharmacy received our alert.

And we also did a survey of the pharmacists' perceptions of the messages that they were receiving, and while the results were admittedly a little bit mixed, they were generally tending towards positive.

We have had two new initiatives that have come out of the work that we've done so far with drug name confusion. One is a potential solution for post-marketing surveillance and risk management. In a manner similar to that that we use today for sending alerts in real time to pharmacies with dose-based rules, we could send alerts for an identified pair that is of particular interest or is a particular problem with other types of rules. For example, if there is confusion between an antipsychotic agent and an allergy agent, we could have a rule around prescriber specialties such that if the antipsychotic were prescribed by an allergy immunologist, that would immediately result in an alert, whereas if the allergy drug were prescribed by that same physician, there would be no alert.

We can also design a method whereby we can send messages randomly. It would completely overwhelm a pharmacy if you sent an alert on every single prescription

for a frequently prescribed drug, but if we can randomly select the prescriptions for that drug that we send messages on, we can still be getting the message out there without having the pharmacist ignore all the messages because they expect to get one.

Retail pharmacies are interested in this service because they benefit by having errors prevented, but they're more interested if they don't have to pay for it.

Also, on the pre-marketing side, we have designed a method by which we can test proposed drug names in tens of thousands -- well, at least thousands. I don't know about tens of -- pharmacies based on the fact that we are connected to 90 percent of pharmacies in the U.S. It's a real pharmacy, so you'll be testing the name in an actual practice environment. You'll be testing it in context with proposed strengths, and there would even be the possibility to try multiple proposed strengths to test the likelihood of confusion in conjunction with the strength.

In many ways it's similar to the methods that Drs. Schell and Hennessy were proposing. I believe that ours could be a little bit lower cost because it's almost completely automated. There is one safety issue that we don't have. We would not propose putting actual bottles of a fake drug or a placebo on the shelf. We think that the

pharmacist just by seeing the prescription and whether or not they can interpret would be enough.

We, of course, would send out prescriptions from fake physicians and fake prescribers and follow up on every single one. So we think there's absolutely no chance that a prescription could be filled on a real patient and take the wrong drug.

And we can compare the results to baseline. We would perform a baseline analysis so you could compare the percent that were cleanly caught and identified as a nonexistent drug, the percent that require clarification, and the percent that are actually interpreted as an existing drug, and compare those to baseline. Of course, in the case where a clarification is required or whether they interpret it incorrectly, we'd be able to tell what exactly it was confused with.

Again, retail pharmacies are interested in participating in this, and they actually see the Hawthorne effect as being a good thing, even though it would be a confounding variable from the name confusion detection side, because their perception is that if the pharmacies know they're being monitored, they are more likely to have better performance at all times, which is beneficial to the pharmacy.

And in reality it would only take three to four

metropolitan statistical areas or the three to four largest chains, and you've got 10,000 pharmacies right there. So it's not an unachievable number.

Of course, the next frontier -- that only covers retail pharmacy -- would be hospital and then electronic prescribing. I think there are possibilities for electronic prescribing, for prescription writing systems. I haven't come up with a solution there yet. And one of the issues there is that the physician initiates the prescription, so there's not anything to react to. So I'm still working on that one.

Thank you for your time.

DR. GROSS: Thank you very much.

Last but not least, Dr. Bruce Lambert from the University of Illinois College of Pharmacy in Chicago. Dr. Lambert.

DR. LAMBERT: I thought that you had forgotten about me.

Thank you for the opportunity to address the committee. Because I only have a short period of time and because I addressed many of these same issues in my public comments during the June 26th meeting, I'd like to direct the committee's attention to my previous testimony and PowerPoint presentation, both of which are available on the FDA website or from me directly or in your briefing

materials.

In addition, I've submitted to the committee reprints of several peer-reviewed articles published by my colleagues and me during the past seven years or so. Although it's not possible to summary the main findings of those articles in the time allotted, each article presents evidence that's directly relevant to the questions being debated today. In fact, they are, to the best of my knowledge, the only peer-reviewed studies that provide evidence as to the validity of computer-based methods for drug name screening.

In fact, many of the questions and issues that have come up today have led to the conclusion that we just don't know about X. And in many of those cases, I was shaking my head because the X that we presumably just don't know about was often described in one of these peer-reviewed publications, especially the relationship between computerized measures of similarity and performance results on behavioral tests of confusion and short-term memory, visual perception, and so on.

I want to talk a lot now about the process of validation for accepting new tests by a regulatory agency. To paraphrase a cliche from the domain of real estate, when it comes to regulatory acceptance of new test methods, there are only three issues to be concerned about and they

are: validation, validation, and validation.

Before a new testing method can be accepted by a regulatory agency, it must be scientifically validated. Validation alone is not enough to warrant regulatory acceptance, but without validation, acceptance ought to be out of the question.

As I prepared these remarks, it occurred to me the regulatory agencies must constantly need to evaluate new testing methods. I felt certain that there would be standard methods for establishing the validity of newly developed testing methods, but I was both right and wrong about this.

On the one hand, there are no uniform policies for the validation and regulatory acceptance of new testing methods across government agencies. EPA, FDA, USDA, NIOSH, and others each have their own approaches.

On the other hand, recognizing this lack of coordination within the U.S. and internationally, toxicologists and regulators from around the world have worked over the last decade to develop a standard approach to the validation and regulatory acceptance of new testing methods. The ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods -- I know that's a mouthful -- also known as the ICCVAM, is a U.S. governmental body run out of the National Institute for

Environmental Health Sciences. Together with a similar group in Europe and from the OECD, the ICCVAM has developed clear guidelines for validation and regulatory acceptance of new tests. These guidelines were developed in the context of traditional toxicology with a special focus on finding new alternatives to animal testing.

But the overall framework should apply more generally to all validation and regulatory acceptance situations. I strongly encourage the committee, the audience, the agency to study these guidelines. They're easily available on the web. Just do a Google search on ICCVAM, you should find them.

It's my recommendation that these guidelines be followed in validating and determining the acceptability of new tests on the confusability of drug names. If they are not accepted, I would request that the agency spell out its own guidelines for validation and regulatory acceptance, and I would also request the agency's rationale for not adopting an existing framework that has proved to be successful elsewhere and is also widely used within the U.S. government.

I want to summarize briefly some of the ICCVAM's main criteria for validation.

First, they define validation as a scientific process designed to characterize the operational

characteristics, advantages and limitations of test method, and to determine its reliability and relevance.

The criteria briefly are as follows. Now, some of them apply, obviously, to toxicology, so some of the vocabulary would have to be modified slightly to think about what are really errors in cognition, for the most part, in the context of drug names. But I'll briefly go over them.

One, the scientific and regulatory rationale for the test method, including a clear statement of its proposed use, should be available.

Two, the relationship of the test methods endpoints to the effective interest must be described.

Three, a detailed protocol for the test method must be available and should include a description of the materials needed, description of what is measured and how it's measured, acceptable test performance criteria, a description of how data will be analyzed, and a description of the known limitations of the test, including a description of the classes of materials of the test you can and cannot accurately assess.

Next, the extent of within-test variability and reproducibility of the test within and among different laboratories.

Also, the test method's performance must have

been demonstrated using reference names representative of the types of names to which the test method would be applied and should include both known positive and known negative confusing names in this context.

These test names should be tested under blinded conditions, if at all possible.

Sufficient data should be provided to permit a comparison of the performance of a proposed new test with the test it's designed to replace. In this case the expert panel is the de facto method.

The limitations of the method must be described. For example -- that's self-explanatory. It goes into more about toxicity testing here.

Ideally all data supporting the validity of a test method should be obtained and reported in accordance with good laboratory practices, which is just sound scientific documentation.

All data supporting the assessment of the validity of the test method must be available for review.

Detailed protocols should be readily available in the public domain.

The methods and results should be published or submitted for publication in an independent peer-reviewed publication.

The methodology and results should have been

subjected to independent scientific review.

So those are the criteria for validation.

They also talk about once a test is validated, how should a regulatory agency determine whether they should accept the validated test because just because it's validated doesn't mean it really fits or meets all the needs of the regulatory agency. So briefly some of the criteria for regulatory acceptance established by this committee.

The method should have undergone independent scientific peer review by disinterested persons who are experts in the field, knowledgeable in the method, and financially unencumbered by the outcome of the evaluation.

Two, there should be a detailed protocol with standard operating procedures, list of operating characteristics, and criteria for judging test performance and results.

Three, data generated by the method should adequately measure or predict the endpoint of interest and demonstrate a linkage between either the new test and existing test or the new test and effects on the target population.

The method should generate data useful for risk assessment, for hazard identification, for dose-response adjustment, for exposure assessment, et cetera.

The specific strengths and limitations of the test must be clearly identified and described.

The test method must be robust. It should be time and cost effective. It should be one that can be harmonized with similar requirements of other agencies. It should be suitable for international acceptance and so on.

So I think these are sound criteria. The report is actually a very, very illuminating one for questions about validation and regulatory acceptance of new tests.

I believe these criteria are sensible and represent the consensus of an international group of experts. They also have some status as policy within the U.S. federal government, although individual agencies are not bound by them. Again, I recommend they be adopted in this context, and if they're not, I request the agency's own criteria for validation and regulatory acceptance be published.

It's worth noting, I think, that none of the methods discussed here today -- none of the methods, including my own, of which I am very proud, but I acknowledge that none of the methods discussed here today meet all of these criteria. I would argue that the methods described by myself and my colleagues come closest, as evidenced by the extensive validation studies published in

peer-reviewed journals.

The methods described this morning by Dr. Dorr and currently being used by the FDA are likely to be sound in my judgment, but they have not been validated in peer-reviewed journals. To my knowledge, there's not a single peer-reviewed publication providing evidence of the validity of the tests being adopted by the FDA, the so-called POCA method. Nor have the operational details of these methods been fully disclosed, and this would violate the criteria for validation as previously described.

I recommend that no method be accepted for regulatory use until it's adequately validated in accordance with the criteria set out above.

So that's generally the issues about validation and regulatory acceptance.

Now I want to touch on a sort of miscellaneous set of issues that have been raised today where I think I might have something useful to add.

The first has to do with the lack of a gold standard. There are many respects in which we lack the gold standard if we're talking about name confusion, and in order to do any sort of validation testing, we obviously need a gold standard.

In one respect we do know what the gold standard is for measuring medication errors and that is

direct observation of real-world medication orders, dispensing, and administration. This is a method pioneered by Ken Barker at Auburn University and generally is the method recognized to be the gold standard method for detecting medication errors. Again, direct observation of real-world behavior. It's the strongest in terms of ecological validity. It's obviously expensive and time-consuming.

There are a variety of other methods which have been discussed today, and I'm generally in agreement with the sort of continuum of having experimental control at one end in the sorts of laboratory tests that I've done and having real-world ecological validity if you do direct observation.

But another sense in which there are no gold standards has to do with the USP list. Now, in my own early publications I used the USP list. I sort of didn't know any better at the time and it was the only evidence that I was aware of. But there are very, very serious problems with the USP list, and in no way should it be viewed as a gold standard. In fact, I think it should be viewed as what I will call an iron pyrite standard. For the geologists in the room, the other word for iron pyrite is fool's gold. So it's the fool's gold standard, and it is so not because the people who are use it are fools, but

because it fools us into thinking it's a gold standard.

So, for example, some names appearing in reporting databases are near misses and not actual errors. So they're status as true positives, as gold standard, truly confusing names is in doubt.

But much more importantly, names not appearing in the reporting databases may, in fact, have been involved in multiple errors but never have been reported. In this case, as Donald Rumsfeld says, absence of evidence is not evidence of absence. Just because a name doesn't appear in a reporting database does not mean and does not even come close to meaning that that name hasn't been involved in an error. Ken Barker's studies comparing direct observation -- and the same is true with Bates and Leape's famous studies of medication errors where they compared direct observation to spontaneous voluntary reporting -- indicate that direct observation yields between 100 and 1,000 times more errors than spontaneous reporting. So what we have in the USP list is sort of the tip of the tip of the iceberg.

This is highly problematic because if we use the USP list as a gold standard and let's say we identify a pair of names that isn't on the USP list, we're going to call that a false positive, but in fact there's no real good justification for calling it a false positive. In fact, it may have been involved in an error that was never

reported.

Similarly, if we say the name that is on the list is not an error, we can't be certain that this is a false negative either because of the dubious status of the names that appear on these lists.

Related to this is the need in any sort of validation testing for the proportion of truly confusing names and non-confusing names to match the proportion in the real world. The problem is we don't know what the proportion of truly confusing names to non-confusing names in the real world. But evaluations of predictive tests, things like sensitivity, specificity, positive predictive value, and so on, which are technical characteristics of a predictive test, all depend crucially on the proportion of truly confusing and non-confusing names in the population.

Next, we're looking at the wrong unit of analysis a lot of the time, and again, I take some of the blame because I myself I think used the wrong unit of analysis in some of my early work. Much of the work on computer methods for name screening, including my own early work, has focused on pairs of names. Clearly there's a certain relevance in thinking about pairs of names because pairs of names are what get confused. But FDA or any other screening agency must approve single names, not pairs of names. So whatever criteria or screening method we use

must evaluate single names, not pairs of names. Methods are needed, therefore, that use the single name as the unit of analysis, not the pair of names. And there are lots of technical reasons why this is so. I'll try to describe just a couple of them.

Any method based on pairs of names will almost necessarily have poor positive predictive value because the sheer number of pairs will overwhelm the false positive rate of the predictive test. That is, let's say you have 1,000 names in the lexicon. Well, there are roughly 500,000 pairs that you get from 1,000 names. If you have n names, there are n times n minus 1 over 2 pairs of names. So for 35,000 or however many trademark names there are, you have tens of millions of pairs of names. Any false positive rate above a tiny fractional false positive rate will totally overwhelm a system if you have that many pairs of names.

In addition, there's this problem that's related to the pair is the wrong unit of analysis but also has to do with frequency. Not nearly enough attention has been paid to frequency. Frequency is a fundamental mechanism of human error, but is absent from most of the discussion about name confusion until very recently, including in my own work until recently. There's been too much focus on similarity.

But the problem is this. All the similarity measures that have been discussed today are symmetric. That is, the similarity between name A and name B is exactly equal to the similarity between name B and name A. The problem is errors are not symmetric. If you have a common name and a rare name that are similar to one another, when presented with the rare name, it's very likely that you will see the common name, but when presented with the common name, it's very unlikely that you will claim to see the rare name. So error patterns are driven by frequency, not just similarity. In fact, in my experiments and in a wealth of psycholinguistic literature, the frequency effect is at least an order of magnitude more powerful than the similarity effects.

So we need to start building prescribing frequency into our predictive models. This recommendation alone is not trivial because there are multiple measures of frequency from the government, from something like the NAMCS database, from IMS, from Solutient. They don't all agree with one another, and so even including prescribing frequency could be complicated, not to mention we don't know the prescribing frequency of a compound before it's marketed, although we have some indication.

We have to think a lot more about non-name attributes. I'm in agreement with a lot of previous

speakers who acknowledged that non-name attributes -- namely, strength, dosage form, route of administration, schedule, color, shape, storage circumstances, et cetera -- are important contributors to errors. The exact magnitude of their contribution is unknown and needs to the focus of future research.

There is the issue of conflict of interest. A lot of money is at stake in naming decisions, both in the naming companies and obviously the PhRMA sponsors. We need to make sure that those doing the safety screening do not have a vested interest in the outcome of the screening. For example, if people who coin the names also do the safety screening, they would obviously have some interest in finding that the name was safe. It doesn't preclude those companies from doing that screening, I should say. They just need to have some safeguards in place.

There's this issue of public costs and private benefits, which I brought up in June. Normally the FDA weighs risks and benefits in drug approval decisions, but here it's difficult to see how the agency would weigh risks and benefits since all the risks accrue to the public, all the benefits tend to accrue to the sponsor of the product.

Harm reduction I agree is the ultimate goal. When evaluating a proposed name, we need to think not just about the probability of error, but about the magnitude of

harm. Harm, as others have suggested is a complex function of the probability of error, the number of opportunities for error, the severity of each error, the probability of not detecting the error, and so on and so forth. Each of these components is difficult to understand because the extent of harm depends on the patient status, the duration of exposure, the duration without the intended medication, the concomitant medications, and so on and so forth.

Just a matter of scope -- I said this on June 26th, but it's worth repeating. The best estimate which we have of the actual number of name confusions in the United States comes from a recent article by Flynn, Barker and Carnahan in the Journal of the American Pharmacists Association, and based on a direct observational study, they report that the wrong drug error rate is .13 percent. That is, they detected 6 wrong drug errors out of 4,481 observations. If you extend that to the 3 billion outpatient prescriptions that are filled per year, that's about 3.9 million wrong drug errors per year, or about 65 per pharmacy annually or about 1 per week in every pharmacy in the United States.

Finally, I want to agree with Maury Tepper and others. I agree with a lot of what Maury said, and I don't just mean the part about being a dumb lawyer.

(Laughter.)

DR. LAMBERT: It's not all about names. Even if we could figure out a perfect screening method for new names, which we will not be able to do, I'm in total agreement this is probabalistic. In the end, the decision will be made by a panel of experts much like this one just like in the end the decision to approve new chemical entities is made by a panel of experts. In spite of the thousands of pages of objective clinical trial data, preclinical trial data, the decision to approve a drug is eventually made by a panel of human experts. That's the way it's going to be here, and it's made on a probabilistic basis. That's the best we're ever going to be able to do.

But even if we could perfect the approval of new names, we would still be stuck with the thousands of names that we have, many of which seem to play a role in confusion. So what are we to do about those?

Here I don't think there's any better authority than Mike Cohen or the people at the Institute for Safe Medication Practices who for years have been advocating safe prescribing practices, safe medication practices, which will minimize these errors regardless of the confusability of names, things like putting the indication on the prescription, dramatically restricting verbal orders, dramatically restricting handwritten orders, using computerized physician order entry, and so on and so forth.

So I add my voice to those who said there's a lot we can do about name confusion other than getting better and better predictive methods for knowing which new names will be confused. While obviously I've devoted a lot of my own time and effort to doing this prediction of new name screening, there's a lot we can try to do to make the system safer and more robust against confusion even with the trademarks we've already got.

Thank you very much for your attention.

DR. GROSS: Thank you very much, Dr. Lambert. These have been excellent presentations.

There was supposed to be one other presenter, Patricia Penyak, who unfortunately was in a car accident and is unable to be here, but her material that she was going to present is in our handouts. So we wish her well.

Is there anyone else who wishes to comment during the period of public comment?

(No response.)

DR. GROSS: If not, let's move on to Dr. Seligman who will tell us the questions they would like us to consider.

DR. SELIGMAN: First, let me thank both the presenters this afternoon as well as this morning for, I think, excellent and thoughtful presentations that I think in many ways have really outlined the complexity of this

topic and really set the stage for what I hope will be a very informative discussion this afternoon.

We have taken the liberty of posing five questions or broad areas that we would like our advisory committee to deal with this afternoon. The first one deals with describing the advantages and disadvantages of evaluating every proprietary drug name for potential confusion versus taking a more selective risk-based approach, considering as we've heard this morning, issues related to consequences, probability, disutility, et cetera, and whether indeed it's possible to develop an approach which would allow us to triage drug names into groups that may be handled differently based on these potential risks.

The second question deals again with many of the study methods that were presented today in asking the advisory committee to give us an assessment of those design elements of those methods that should be included in a good naming practices guidance and what elements of those methods should either be discounted or not considered useful in developing such guidance.

Third, we would certainly like to hear from the committee if there are, indeed, other methods that should be considered in producing such good naming practices.

Finally, we'd be very interested in learning

under what circumstances field testing in a simulated prescribing environment should be considered. I think it's pretty clear, based on what we've heard today, that it's unlikely that one method alone would be sufficient, and clearly we're interested in learning what combination of methods should be deployed such as behavioral testing and orthographic and phonographic testing or other combinations of methods.

Finally, we'd be interested in hearing from the committee as to whether there are circumstances, if any, when it might be appropriate to approve a proprietary drug name contingent on either some element of a risk management program being in place in the post-marketing environment.

With that, Mr. Chairman, I turn the discussion to you.

DR. GROSS: Dr. Seligman, could you clarify the last question? When you say approve a proprietary drug name contingent on risk management program, that means that for some reason the name will stick rather than trying to change it or because the drug is risky and you want to have a risk management program?

DR. SELIGMAN: No. It's basically essentially allowing a name to be used knowing that there might be a potential for, I guess, confusion and the degree to which one might want to more carefully assess in the post-

marketing environment indeed whether harm occurred as a result of allowing that name to proceed into the post-marketing environment. Jerry, is that the interpretation?

MR. PHILLIPS: Yes.

DR. GROSS: Okay, fine. Thank you.

Is it the committee's pleasure to do this one at a time starting with number one? Okay. Does anyone want to comment on number one? Advantages and disadvantages of evaluating every proprietary drug name versus taking a more limited approach based on risk.

MS. JAIN: Well, Dr. Gross, I just want to say that you had mentioned previously that you wanted the FDA representatives and the PhRMA representative, Mr. Lee, to produce lists of how they do their analysis in a step method. I distributed the FDA version that Jerry Phillips was nice enough to write up, and I've got copies for the committee members from Mr. Lee as well that I'll distribute at this time.

DR. GROSS: Okay, good.

Brian.

DR. STROM: The question is whether all drugs should be screened or whether a risk approach should be used. My sense is that all drugs have to be screened because even if the drug itself is a low-risk drug, you don't know which drugs it's going to be confused with.

They, in turn, may be high-risk drugs.

I think the place that the level of risk would come into play is more related to the fifth question, that if in fact the therapeutic ratio of both drugs is low so that they're both relatively safe drugs, you might be more willing to tolerate allowing a drug name on the market despite the risk of confusion. So your threshold for a decision may be different, but it's hard to imagine you could not screen all names given you don't know which drugs they're going to be confused with.

DR. GROSS: I see a lot of nodding heads on Brian's response. Yes, Curt.

DR. FURBERG: Yes, I agree with Brian. I can see a step-wise approach. You start off with screening, probably very simple or simplistic.

The issue really is how do you define a high-risk drug. That is the crux. Where do you draw the line? I'm not sure I know exactly how to take a stand on that. But clearly, step-wise makes a lot of sense.

DR. GROSS: So that's the second part of the question, but for the first part, does anybody disagree that all drugs should not be run through an approach? Robyn.

MS. SHAPIRO: I don't think I disagree. I just want to be sure that I'm understanding this right, and that

is that at the moment this happens in two different spheres. One is the FDA already does that. That's the practice now, and two, the whole trademark process, as we heard about, also is a way of screening for this very thing. Is that right?

DR. GROSS: No. I think that's a separate issue.

MS. SHAPIRO: Okay.

DR. GROSS: We're not saying who's going to do the screening. Right? Is that your question?

MS. SHAPIRO: No, no.

DR. GROSS: Paul, is your question whether the FDA should do the screening or somebody should do the screening?

DR. SELIGMAN: No, it's not a question of who. It's a question of whether, whether it should be done.

DR. GROSS: Right. That's what I assume. Okay.

MS. SHAPIRO: And I'm just trying to confirm on the whether, not the who, that there are two systems already in place doing that.

DR. GROSS: Okay. That does not happen to be one of the questions of the five, but it's certainly something that we can comment on because it's an issue that's come up over and over again. If you want to discuss

that -- you know what? Why don't we go through the questions here and then come back to that particular point because it is an important issue.

MS. SHAPIRO: Okay.

DR. GROSS: So it sounds as though everyone agrees that all proprietary drug names should be screened. We're not specifying how.

Yes, Stephanie.

DR. CRAWFORD: Thank you. Just to clarify our recommendation, would this be every drug name screened pre-approval? We're not talking about retrospectively looking at all existing proprietary names?

DR. SELIGMAN: That's correct. Pre-approval.

DR. GROSS: Yes, Lou.

DR. MORRIS: Does that include OTCs on switches?

MR. PHILLIPS: Yes.

DR. MORRIS: Are they screened now?

MR. PHILLIPS: If they are subject of an application, they are screened.

DR. MORRIS: So if a well-known prescription drug that's on the market is switched and has the same name, it has to go through new testing?

MR. PHILLIPS: It usually has a modifier or something associated with that trade name and it will go

through an assessment.

DR. MORRIS: Oh, okay.

DR. GROSS: The second part -- yes, Jeff.

MR. BLOOM: I just wanted to add one thing. I agree with that as well. I'll just add to the point that even a drug that seemingly may be innocuous, we have to recognize that many drugs are used in combination, and whereas a drug may seem to be rather safe, but when used in combination might have some other side effects or interactions, I think it's very important that it all be screened. I agree completely that it should be screened ahead of time.

DR. GROSS: How about the second part of question number one? Is it possible to triage the drug names into groups that may be handled differently based on risk? So an initial approach is a yes or a no, and if yes, how? Eric.

DR. HOLMBOE: I think in fact what Brian said earlier, it would be difficult to do that until you know what it's look-alike actually is. If it turns out it's a low-risk drug, but it's similar to a high-risk drug, then it's hard to triage based on the single agent.

DR. GROSS: Yes, I agree too.

Does anybody else want to comment on that part? Arthur.

DR. LEVIN: A point of clarification. There are several risks here. One is risk of confusion, one is the risk of toxicity. And there are probably a lot. We can make a long list of risks, so we just need to be clear when we talk about potential risks that we agree what we're talking about.

DR. GROSS: Paul or Jerry, do you want to comment on that?

DR. SELIGMAN: When we talk about risk, we're pretty much talking about risks of adverse events, basically the consequences, the probability, the disutility, some of the things that Sean Hennessy addressed this morning.

DR. GROSS: So it sounds as though the answer is no to the second question. Anyone else want to comment? Lou.

DR. MORRIS: Is it possible? The answer is yes. But is it advisable is the question. Clearly you can put drugs in categories based on the severity of the adverse event, but I think the question here is is it advisable to do that, and I don't know the answer.

DR. GROSS: Fair enough.

DR. STROM: Yes. To just be clear, I completely agree with that. It's possible to stratify based on the risk of the error with the parent drug, but

we're saying that in initial screening you shouldn't do that because it's impossible to know what the risk is of the drug it's going to be confused with because you don't know yet what drug it's going to be confused with.

DR. GROSS: The second question then is based on discussion of the study methods presented today, identify the critical design elements of each method that should be included in good naming practices. I'm not clear on that question. I mean, we're not really going to discuss the critical design elements in each of the methods. Is that what you want us to address? Or did you want us to say what study methods should be used in trying to avoid confusion or what combination of study methods?

DR. SELIGMAN: I think either what methods or what combination of methods, but also particularly within some of those methods, were there elements of them that were particularly strong or important that should be emphasized in constructing good naming practices?

DR. GROSS: Yes. I think Dr. Lambert made a very good point that there are very few that have been validated except for the ones that he described. If anybody disagrees with that and is aware of other validations, please speak up.

So does anyone want to comment on that first sentence? Brian.

DR. STROM: I wanted to make a number of comments. I've been writing notes and this seems to be the appropriate question to respond.

I think what we heard today and in June is striking, that in a sense in drug names, we're equivalent to a pre-FDA era in drugs. It's as if we were approving drugs based on preclinical data only and no clinical data. We're approving drug names here based on data that has never been validated, and we don't know what the interpretation of any of it is.

We hear, on the one hand, that industry thinks it's a tiny problem. We hear, on the other hand, FDA rejects a third of the ones that industry thought were a non-problem. And we don't know which one is right based on the available information.

We've heard many people talk about their best practices and everybody should use best practices, but none of those best practices have been validated to know that any of them are in fact best practices. A lot of cutting-edge, very exciting new methods that we're hearing about -- and I'm very interested and excited by all that, but none of that has yet been evaluated.

So I guess my own biases would be, on one hand, to be careful. I would not recommend changing a current process, given we don't know what's right and what's wrong

with the current process. But I would recommend we don't know what's right and what's wrong with the current process and we need an enormous amount of work very quickly to do the needed validations and to use simulations and laboratory techniques and the kind of thing Sean talked about and whatever as ways of trying to find out what works and what doesn't. We probably shouldn't change much until then because, again, we don't know that there's a major problem out there. The current system with industry doing it and then FDA doing it may well be fine, or at least, parts of it may well be fine and you don't want to risk throwing out parts that work, given we don't know what works and what doesn't work.

DR. GROSS: Other comments? Michael.

DR. COHEN: I also jotted down some notes.

I think the expert panels, the focus groups are important, and that is current practice I think for most of the companies. I think it picks up the kinds of things that some of the other testing may not. For example, the computerized systems that we heard about today would not pick up some of the prescribing-related problems like stemming of a drug name, those kinds of issues that sometimes cause confusion with a drug that's already available.

I think also the value of the nurses' input and

unit clerk input and pharmacists' input is immeasurable. True. But I think it's very important. They're likely to pick all kinds of things: confusion with prescription abbreviations, for example, parts of a name that might be confused with a dosage form or the dose or quantity, as we heard. So I'd like to see that continue.

The computer matching. I could see that being used in conjunction with it. I mean, it is a validated process. We've heard that. I think it depends largely on the type of database that's used, what the database is. For example, there are some databases that contain names that are not really drugs on the market, and you'll get printouts of that. I also --

DR. GROSS: Michael, I thought it was said that the computerized systems have not been validated.

DR. COHEN: I thought that Bruce said that it was. His system. Did I miss that?

DR. LAMBERT: Am I allowed to speak?

DR. GROSS: Yes. Bruce, do you want to comment?

DR. LAMBERT: The methods that I propose and have been working on for the last seven or eight years have been subject to extensive validation testing. This is not to say they're perfectly valid. When you subject a method to extensive validation testing, what you find are both its

strengths and its limitations. What I argued was that the methods that I have described are to my knowledge the only methods for which there are peer-reviewed articles about the status of their validity.

DR. GROSS: Yes, I know. Bruce, Bruce --

DR. LAMBERT: And certainly my methods, I validated them against visual perception, several different short-term memory tests, against the perceptions of established experts, against the perceptions of lay people, against databases of known errors, and so on.

So the methods that I propose, the bigram, trigram, Edit, et cetera, are by no means perfect, but I have documented in extensive detail the extent to which they are valid. Those materials are in your briefing packets. I sent them to the agency weeks ago, but I'm told that you only received them today. So if you haven't read them, I understand. They're not exactly as exciting as a John Grisham novel. But these methods have been subjected to extensive validation testing. It's up to your own judgment as to whether you think they are valid enough for use for these purposes.

DR. COHEN: I want to point out that I don't think they can be used alone without any doubt. I think they can be used in combination.

DR. LAMBERT: And neither do I. In all of my

publications, I say they shouldn't be used alone.

DR. GROSS: Bonnie.

DR. LAMBERT: I say they should be an input to an expert process.

DR. GROSS: Bonnie.

DR. DORR: I just wanted to point out that there is currently under peer review an article on an evaluation of different techniques. One of them is ALINE. Another is -- as I mentioned this morning, our best result was a combination of ALINE with a bunch of other techniques where we're getting high results with the caveats already mentioned in my talk and also Bruce Lambert mentioned that the data that you have as a gold standard -- we're having problems with that. We're using USP. We did use a smaller list of known error drug names that are not the USP list also, and we were getting similar results.

And the technique itself of ALINE, outside of the task of drug name matching, has indeed been validated by several peer-reviewed articles. There's a Ph.D. thesis on it but, again, that wasn't for the task of drug name matching. Right now, within two to three weeks, we should know the answer for a particular peer-reviewed article for this task, and we'd like to talk more about the combination of different approaches and also not just within the computerized technique, but outside of that. What can we

combine those computerized techniques with to get what you need.

DR. GROSS: Right. That's a separate issue.

DR. DORR: Because as Bruce said, you can't just say it's valid for this test. Even if you say the algorithms are, indeed, measurable up against each other, it may not be appropriate for this task.

DR. GROSS: Thank you both for the clarification.

Michael, do you want to continue?

DR. COHEN: Yes. Let me continue.

Where I think it can be valuable is if something might be overlooked with the review by practitioners, the group testing, et cetera, I think that that can help as kind of a backup system that further assures that something important is not overlooked. So that's why I see this being used only in combination, not by itself.

Then thirdly, about the model pharmacy and the laboratory. I can definitely see where that could be helpful post-marketing. Pre-marketing, at least at this time, until we see some evidence of its value, I could see a lot of problems with it, and I don't think that that would be of value at this time anyway until we see it actually proved for the reviews.

DR. GROSS: Curt?

DR. FURBERG: Well, it's clear that we have multiple methods. They all have strengths and weaknesses, and so I agree with the idea that you need to somehow develop a battery.

My sense is that people in the field are not communicating very well, and there seems to be some turf issues also. We can't settle that in a hearing.

So my suggestion is that the FDA appoints a working group of all the experts and let them come up with a recommendation of an appropriate battery that could be discussed, come back to the committee, and then we can move forward.

DR. GROSS: Ruth.

DR. DAY: The problem that we're having right now is there are several different methods and each have several different design features. Each design feature has advantages and disadvantages. So if we had the list before us and we had a lot of time, we could do that, and maybe Curt's suggestion would be good.

But if we were to go down each element in each method, it could be very useful. For example, an expert panel. In round one, as I understand it, people independently generate sound-alike or look-alike candidates for a given drug name. Well, where do those come from? So

some of the people might just take it out of their heads, out of memory, availability in memory. Some might go check the PDR. Some might look at the USP database and so on and so forth. You want people to be able to do whatever they do because that's what they're going to do in every day life. But you could document it a bit. So for each focus group, after it's over or after round one is over, you could get that information.

So a big problem in all of this is noise in the data and lack of replicability. And it could be that by getting more information like this, you could say, oh, focus group 1 all looked up in the PDR. Focus group 2 had a mix of other methods to generate and so on and so forth.

So especially for whatever is the first step in all these processes, such as generating potential names to consider -- that might be difficult -- or in the case of the linguistic methods, there are other things to do first like pronounceability, which I'll comment more on later.

DR. GROSS: Yes, Eric.

DR. HOLMBOE: Also, I just want to highlight that it was my understanding at the beginning that your hope was that in time industry actually would take a greater responsibility for this. And so far, I think what we've talked about is actually what you're doing. Clearly the strengths and weaknesses due to that and I think we'd

all agree that a multi-factorial approach is probably the best.

But I would be interested to know actually what industry is doing. We haven't heard a lot about that. We didn't get a lot of data, but clearly there's a big disconnect. We've heard from several groups today that they feel that they're doing a fair amount in this kind of pre-marketing work, and yet, as we heard, you reject a third of the names despite the amount of effort that they're using to try to come up with a drug name even before it reaches your desk, so to speak. So I think there needs to be a better understanding of why we're seeing such a disconnect, particularly if we're going to migrate the methods back into the private sector for them to take care of it instead of you doing the things you currently do.

The second thing I would highlight is that we've heard from the epidemiologic perspective that what you're really trying to look for here is a really good screening test. So you're really looking for something that's going to give you high sensitivity, and then how do you deal with the kind of false positive rate that gets generated out of that? Clearly that's another issue that we haven't really brought up today, but in a sense that's what we're talking about with a lot of these things that we're really trying to screen. So that would be another

principle.

The finally, I'd encourage you to look at the Medical Research Counsel out of Britain actually which has done a very nice monograph on how to approach conflicts intervention. That's what you've got here. You've got multiple methods that you're using. And they provide a very nice framework to think about how to move this forward over time that perhaps the working group would be able to use as well.

DR. GROSS: I wonder if Paul or Jerry might comment on why the high rejection rate on the names from industry when they've gone through the screening that they have told us. They've told us they have gone through most of the screening methods that have been described.

DR. SELIGMAN: I don't know the answer for sure, but I'm happy to speculate because I suspect that there's probably a wide diversity within industry as to the kinds of techniques that they've applied. I think what you heard today, if I again would venture to speculate, is probably the best practices that probably are, indeed, well conducted by many of the major pharmaceutical companies.

I don't know, Jerry, whether we have any analyses that we've done on looking at those we've rejected and whether there's any difference by company size or generics versus proprietary names or whether there are

clues as to why there seems to be that disconnect.

DR. GROSS: Yes. I see Bob Lee's hand is raised. We were going to ask him, even if he didn't raise his hand, to comment.

MR. LEE: I thought it might be helpful to just explain what it is we do do as part of our screening. A lot of it initially is what you'd really call data acquisition. Well, even before that, first we have to generate new names. They have to be created. We can do that in-house. Anybody can sit down and come up with coined or arbitrary names. These are names that don't mean anything, but which are pronounceable. But we usually use more expert groups, branding companies who know how to do that a little better, who may have been in the advertising area or have other backgrounds in creativity, if you can define what creativity is.

So they generate long lists of names that then are submitted to the company, usually to a team within the company that's made up of different disciplines. There are so many initially, 100, 200, 300 names, that they have to be narrowed down into a smaller, more manageable group for extensive searching. So some are thrown out just because they're not liked and some obviously have bad connotations or remind people of bad things, or for a variety of different reasons many of those names are just thrown out

from the beginning where people can spot confusion problems immediately upon seeing some names.

But then you get down to a group of names, perhaps 30 that you begin a very extensive searching process on using various algorithms like the algorithms we've seen although maybe not identical to Dice coefficient of the kinds of letter-string systems that we've seen, or the phonetic tools that we've seen today are very powerful. So not necessarily those, but where you will take prefixes, suffixes, letter strings and combine them in various ways to try to pull out of the database that you're searching other names that look similar to the one you think you want to go forward with.

DR. SELIGMAN: Bob, do you know how common these practices are within industry, and can you speculate as to why there seems to be a disconnect between the rejection rate of names within the FDA and your view that, indeed, this work is being done very thoughtfully and carefully within industry?

MR. LEE: Well, I think your point is actually a very good one about whether or not all of the companies who eventually submit names to the FDA are following these practices. I'd have to say I think most of the major PhRMA companies that make up the PhRMA organization are following similar practices. They're not doing everything that we

might list, but they're doing many of them. Almost all of the major PhRMA companies are doing extensive searching in databases using algorithms.

That's not to say that there can't be improved algorithms and certainly improved databases where all of the factors we talked about can be accumulated in that database so that they're readily available to the searchers. That makes a more comprehensive review possible because otherwise you have to do the trademark searching, though names only, and then you have to do investigations about the names that you're seeing that might be confusingly similar to the ones you're going forward with. You then have to do a lot of searching to find out what's the dosage amount, so on and so forth.

Of course, getting information from front-line practitioners about that is very, very helpful, but sometimes it's difficult to acquire that data.

DR. GROSS: Arthur.

DR. LEVIN: Two comments. Paul, with all due respect to PhRMA, I would suggest that the purpose behind trademarking is not primarily safety. Trademarking, one, has a legal aspect that's very powerful, and it has a marketing aspect that's extremely powerful. I don't mean to suggest that the safety is disregarded, but trademarking is not a principle or a concept or an activity that was

developed in the field of safety management, risk management. Number one.

The second thing. In a way, equally interesting to the question of why this disconnect where a third of the names that go through this rigorous process are rejected by FDA is what about the names that FDA accepts. They've gone through a rigorous process by PhRMA, and then they're accepted by the FDA's rigorous process, and then lo and behold, we find significant problems in confusion.

Have we taken a look-back at those failures, so to speak, and said what happened here? How did it get through both of us, and what was missing in our process? Because it seems to me to answer the question about what's needed in terms of what sorts of combinations of processes can best eliminate the problem or reduce the problem is to know where the failure has been. It's like dealing with error and learning from error. We go back and look at what went wrong to discover how to do it right, and I think the same principle should apply here.

DR. GROSS: Yes, doing your own RCA or FMEA.

Before we try to come to some conclusions on question 2, let's take a look at the second sentence in that question. Are there any methods that should be discounted as not being -- and the key word is --

potentially effective. So there are some tools that we've heard have not been validated but potentially they may be worthwhile. Does anyone want to discount any of the methods that we've heard?

DR. MORRIS: I wouldn't discount per se, but I was struck today that I felt certain tools or certain techniques were -- I was comfortable as seeing them as hypothesis-generating techniques, but not confirming, and yet simulations I felt I was more comfortable with at least their potential. So maybe we can separate them into hypothesis-generation techniques and possibly confirming techniques as a means of putting them in some category.

DR. GROSS: Okay.

Yes, Michael.

DR. COHEN: I guess I disagree a little bit with that only because, like I said before, I haven't seen them proved yet, number one, and I know you'd agree with that. Number two, they really do see a little complex and perhaps not so practical to actually carry out for trademark reviews when large numbers of names are being used. They don't include all environments in which the drugs are used. I don't know that they couldn't be set up. All I'm saying is I think it needs a lot more work.

DR. MORRIS: I used the word "potentially" very carefully there because I agree that because they're not

validated or we don't know enough about their validation, I'm not comfortable saying how they should be designed, but I think they have more potential for giving us better data.

DR. COHEN: I would say that they definitely would hold promise, but it needs more work.

DR. GROSS: Yes. I'd like to propose as a possible approach to the whole of question 2 to follow up on what Curt Furberg said and that maybe the FDA could appoint a small group of people to come up with maybe a minimum combination of methods. Does that fit what you're talking about, Curt?

DR. FURBERG: Yes.

DR. GROSS: A minimum combination of methods and then if people want to supplement it with other methods, fine. It's always hard whenever you take a multi-faceted approach and you're picking from a menu of many different methods how to pick which ones will work. There aren't too many studies done in various fields where that's been elucidated.

DR. FURBERG: But I think it's also important to have broad representation. I think PhRMA should be involved, should be represented on that committee.

DR. GROSS: Sure.

DR. STROM: Can I have two comments on that? One is to some degree the June meeting was that in terms of

having groups talk to each other and with each other and communicate.

The second, what's really needed is what you're describing in terms of a work group doing it, but it needs data to work with. The groups, having now talked to each other in June and now presenting here, it's not clear to me that a meeting yet -- I think that kind of meeting is exactly what's needed after there's some data for the meeting to react to because everyone can give an opinion, but it's like saying I think this drug is effective because in my experience it worked before the era of clinical trials. Until we have some scientific data to know what works and what doesn't, all we're going to hear is more opinions and more expressions, best practice, without a basis behind it.

DR. GROSS: So in the absence of enough scientific data, would you like to make another proposal?

DR. STROM: I think there needs to be a major -- well, that's why one of my suggestions before is that I wouldn't change things much now yet in the way things are done. I certainly wouldn't abandon what FDA is doing, in terms of shifting it to industry, given a third of the drugs it's getting from industry it's now rejecting. But I think a major effort is needed for a large research effort in order to generate data evaluating these approaches.

Once those data are available, that's the time to hold the kind of meeting that Curt described.

DR. FURBERG: Yes, but you can't talk about it sort of globally. We need new research direction. Who's going to provide those? You need that expert group to sit down and say this is what we know, this is what we don't know, and then develop a plan from that.

DR. STROM: The people are going to provide it, the researchers. There is no lack of researchers in this country. And if FDA would issue, as a challenge to PhRMA, RFAs to say let's evaluate the methods that are now being used.

DR. FURBERG: I would be more in favor of a coordinated effort rather than what you're talking about, an isolated effort by people who have self-serving interests to some extent and pursuing their own ideas. I think we need to get together. All the parties should be involved. We should discuss what we know and what we don't know and then develop a plan.

DR. GROSS: Any other comments from the committee? Yes, Jeff.

MR. BLOOM: Yes. On the Regulatory Reform Committee, which I was a member of, we did have recommendation 238. The reason to shift doing the safety testing to industry was the recognition of the limited

resources of the FDA frankly, which is part of the problem in this issue. The idea was that to review data from sponsors who followed protocols designed to evaluate potential for look-alike and sound-alike errors with generic and proprietary names prior to FDA-regulated drugs and use the information gathered from that name safety research to improve patient safety. One of the ways you would improve that is looking at Medwatch reports -- you do get adverse events from naming problems and things like that -- and see which ones are minimized and which ones are not. You can look at those protocols and that way you'd have some sort of baseline at least to start looking at some systems that may be potentially beneficial for naming things. The real question is the resources that you have to put into this are quite limited, and that was one of the reasons that we thought that would be a good approach.

DR. GROSS: Jackie.

DR. GARDNER: Along those lines, something that Brian started with today about the gold standard, I think at an absolute minimum -- I'm left at the end of all of this discussion in not really knowing which things are serious, what is the gold standard, which confusions have resulted in harm as opposed to confusion, and it's something that I know PhRMA raises all the time. Is there a risk here?

So I would like to see some targeted work done both in-house and maybe under an RFA about looking at some of the things we've heard about. We heard that the USP gold standard combines both things that have been known to cause harm and things that have been just reported and we're not sure or things that were caught, potential. We heard from Jerry I think that it isn't exactly -- I want to paraphrase, but tell me if I misunderstood what you said. They don't know exactly which of the things they stopped -- they don't have good numbers or a clarification of which things caused harm that were let go through.

So I guess if we could begin to clarify those things as a baseline, there may be patterns buried in there that would help to then direct some of the other work. It may be only things that have four strings are the serious ones. I don't know. But I don't feel that we have that foundation to begin with about what is really potentially harmful.

DR. GROSS: Any other comments? Arthur.

DR. LEVIN: I just want to caution that today's near miss is tomorrow's error. So I'm cautious -- and I think we were in the IOM -- about the relative weighting of things that actually cause harm and things that don't. I think they are different, but just because something gets caught doesn't mean tomorrow it will get caught.

I think the problem with the gold standard, with all due respect to my friend Mike, is that by relying on voluntary reporting, our n's are always far from what we would like them to be and to give us all of the information we should have. This is not a plea for mandatory reporting. I'm just saying it's a fact of life that the voluntary reporting systems have not been nearly as productive as we would have hoped they would be, and I don't know how to address that.

DR. COHEN: You mean in producing numbers.

DR. LEVIN: Yes, in producing numbers.

DR. GROSS: Brian.

DR. STROM: I certainly agree. I think the bigger problem with the spontaneous reporting system, as was described before, much more than the sample size is the selectivity, that you don't know what you're missing and undoubtedly you're missing most of it. Overwhelmingly you're missing most of it. So I'm very, very nervous about using that as a gold standard for that reason.

On the other hand, I certainly agree that near misses could well be important later, but it depends on how you define them. For example, direct observation. People look at these vast numbers of medication errors. Well, some of those medication errors, a large number of them, are things like getting a drug -- if you do direct

observation in the hospital, they list as a medication error getting a drug 15 minutes late. I'm not worried about that as a near miss, and that's not going to be a disaster later for most drugs. So it is still important to look at which of the medication errors matter and which are the ones that don't.

DR. GROSS: I'm going to make a proposal here. In the absence of enough data for us to make firm recommendations, what would you think about recommending sort of a modification of what Curt said, recommending that the FDA meet with PhRMA and decide whether to maintain the status quo until we have more experimental data to make reasonable decisions on or whether a change should be made?

DR. DAY: Can you modify that to say PhRMA and other groups? It's not just a PhRMA issue.

DR. GROSS: Yes, sure. Do you have a particular group in mind?

DR. DAY: All the usual stakeholders are potential candidates.

DR. GROSS: Okay.

Michael.

DR. COHEN: I think we ought to be very careful with that, though, because I want to make sure that nobody walks away with doing nothing. So that needs to be qualified in some way. I think at least what's being done

now is absolutely preventing some potentially dangerous names from getting on the market at all. So to do nothing would be not the right way to go.

DR. GROSS: Wait a minute. Are you saying --

DR. COHEN: You said if things should stay the same, status quo, or not.

DR. GROSS: Right.

DR. COHEN: So I say qualify it by saying you don't want to go back to doing nothing.

DR. GROSS: Well, no, we don't. We're not doing nothing now.

DR. COHEN: Correct, but the way it was stated I think left the impression, at least for me, that one of the decisions could be we would do nothing.

DR. GROSS: No, no. That wasn't what I meant to imply.

Brian.

DR. STROM: Yes. I would suggest a modification of it. I'm not comfortable with the way you worded it in the sense of I don't see how FDA could meet with PhRMA and decide whether or not to make a change, again without any data. Without any data, I don't see there's a reason to make a change. I would suggest that FDA should be meeting with PhRMA and other relevant stakeholders to decide what data are needed in order to

decide and design a plan to gather those data.

DR. FURBERG: And bring it back here.

DR. GROSS: That's fine.

Yes.

DR. CRAWFORD: Thanks. I would like to echo what Brian just said because with the handwriting problems, I had to look a few times.

(Laughter.)

DR. CRAWFORD: I do appreciate the analysis of the processes presented both by the agency and the PhRMA representative. What I didn't see on the FDA steps was interaction with the sponsor. What I didn't see on the sponsor's steps was interaction with the FDA. So I'm wondering as part of the process, at some point if the proposed nomenclature is problematic for FDA, is there a step whereby the FDA interacts with the sponsors and is the sponsor given the opportunity to present safety information, a similar level of validation as you do with all the other benefit-to-risk safety data presented in an application. And if that is not done, then is it just a second-choice name or what happens?

DR. GROSS: Jerry.

MR. PHILLIPS: The process is reconciled at the end of the day when they're given a choice of either coming back with another name or coming back with persuasive

evidence. So a sponsor has the ability to go out and do a study or provide us the data to persuade us to change our opinion. So the sponsor always has that ability to persuade us to change our mind or to submit another name for review.

DR. FURBERG: But, Jerry, before you get to that stage, before you reject it, you need to sit down before the name is submitted almost to agree on the plan how you find out about this name confusion.

DR. GROSS: Yes. I think we could spend the rest of the day and the week debating this issue, and the reason we're debating is because we don't have the data we need to make a reasonable recommendation.

So, Brian, do you want to restate your version of everybody else's version, if you can remember?

(Laughter.)

DR. STROM: I guess my recommendation would be that the current process not be changed on both sides, the FDA or industry, absent data to the contrary, but that we're not affirming that it is the correct process. Our recommendation is that PhRMA, FDA, and all the relevant stakeholders meet to discuss what data are needed in order to, in fact, find out which approaches are correct and to develop a mechanism for generating those data.

DR. GROSS: Okay. I hope nobody wants to amend

that.

(Laughter.)

DR. FURBERG: And bring it back here.

DR. GROSS: And bring it back here. Accepted.

All in favor, raise your hands, please.

(A show of hands.)

DR. GROSS: Thank you. That was a tough one.

The next one hopefully will be a little bit easier. Are there any other methods that were not discussed today that you think should be considered? Ruth?

DR. DAY: I'd like to suggest a method which is quick, easy, cheap, and I think very valuable. It is pronunciation screening in a systematic way. A lot of the methods we've heard about today assume that a drug name has a pronunciation. In fact, drug names often have alternative pronunciations. We've heard today quinine, quinine, quinine. We heard about Novicar, a made-up name. It could also be Novicar. It could be a lot different things. And does it matter? As the old song said, you say Arava, I say Arava, but it doesn't make any difference because we understand each other. That's a case where perhaps it doesn't make a difference.

However, there are many cases where the pronunciations that people give, when they first see a drug name, are wildly different. So for amoxicillin you can get

amoxicillin. For clonazepam, you can get clonazepam, clonazepam, clonazepam, clonazepam, et cetera. You can get wild variations. So how do we know what the effective pairs are to be worrying about in the first place.

So I'm concerned that the horse has gotten out of the barn in a lot of these methods before the appropriate phonetic cart has been attached. We don't know then how --

DR. GROSS: Or that there are a lot of other horses in the barn that we haven't seen yet.

(Laughter.)

DR. DAY: Not only are there other horses in the barn, but we don't know which ones to be comparing. So this can account for the incidence of both false positives and false negatives. So we may be identifying "problem" pairs by linguistic methods, where in fact psycholinguistic methods where people would pronounce in advance would say, no, people aren't going to be confusing those. Also, false negatives where we think a pair is okay, but in fact, the way people pronounce them would make it not an okay pair.

So a very simple task. A person sees a drug name and says it out loud. Of course, you have a bunch of different ones that you present. The main dependent variable is agreement and the different pronunciations that are given, and I'll come back to that in a moment. Also

speed of naming and the number of attempts to repronounce and change one's mind about how it's said. So on the agreement side, a given drug name -- does it only have one pronunciation, and does everybody agree? That would be great. Go ahead. But if it has multiple ones, what is the probability of each one? So if it has two, but one is 95 percent and one is 5 percent, that's different from if you have a 40/40 and then some dribbling off. So the overall frequency distribution of pronunciations can be very informative.

Once you have this set of data, you can then look at the effects on both other cognitive tasks and on behavior. For cognitive tasks, free recall. What were the names of the drugs you just saw? Can people even say them or remember them? Or give a recognition task. Show them one at a time and say is this one of the drugs you just saw or not, and then you can put in potential confusable pairs and so forth.

So very quickly, the advantages and disadvantages of this very quick little thing are the following. The advantages are it can be very quick. You can do an effective experiment or test in even 5 to 10 minutes, depending upon what you include in it and so on. It's easy to do. It's inexpensive. The data are quantitative. They are easy to replicate. The data are

objective. It's easy to understand the results. It's easy to apply them in a variety of ways, and this approach may well reduce the noise in all the data of all these other methods. So when one of the wonderful linguistic analyses that makes great sense from a linguistics and computational standpoint does not identify or has some kind of problem, it might be because of pronunciation alternatives.

Also, with the outcomes of these studies, we can determine pairs are then likely to be confusable, and the probable pairs or likely pairs are likely to change relative to what we have now. And building on something that Bruce Lambert said, this is also a way to evaluate a single drug name before you start looking at any pairs.

Of course, there are limitations. Every method has limitations. It only is addressing the sound-alike problem. It cannot stand alone, obviously. And it's only really for initial screening. But it could be used later on as well as new products start coming on the market and maybe they come in through some route and they're there so that a sponsor could launch a risk management approach based on something that happened. So it could be a TV ad. I say Arava, you say Arava, but together we agree that it works. I don't know. Whatever it would be. But some kind of approach could be taken then to handle things that come up.

On the sponsor's side, you can then reduce that tremendously long list of 100 to 200 to 300 names that you generate right away by looking at the pronunciation data in a systematic way, not in expert groups sitting around and doing it because I think we need to have a variety of different participants in such tasks from the health care professionals, the doctors, pharmacists, nurses, and the lay public, the patients and the caregivers and so on, to see the variety of namings that would happen.

On the linguistic models, they could then perhaps start with more realistic phonetic transcriptions, as Dr. Dorr admitted this morning or acknowledged, but also they might discover new variables that need to be taken into account. I didn't hear anything today about analyses about syllabicity. How many syllables and where are the syllables segmented and the stress and intonation contours of how you say something? So the stressed and louder and higher-pitched syllable is then the one perhaps going to be more likely to be confused with other things.

For regulators, the advantages of having something like this are that they could replicate using the exact same methods within one day on these things, and they could then have standardized methods across all of those people who want to do some kind of testing.

So, in conclusion, whether there is a screening

test or not for pronunciation or pronounceability, it is an essential ingredient in all this and could be responsible for some of the problems across the methods.

DR. GROSS: Ruth, thank you very much. We expect to see the results of your study published in a peer-reviewed journal soon.

(Laughter.)

DR. GROSS: Yes, Lou.

DR. MORRIS: Yes.

DR. GROSS: I think it was a very good suggestion, Ruth.

DR. MORRIS: I'm not totally comfortable that we really understand the root cause of sound-alike/look-alike problems. We're making an assumption that there's a problem in the communication between the doctor and the pharmacist per se.

I was struck with something Jerry presented that there are actually a lot of problems with doctors writing the wrong name, and I think there may be memory retrieval problems that doctors have recalling the wrong name. I guess what I'm suggesting is as part of this research that we're suggesting, as we understand these root causes better, there may need to be different methodologies in the future and that we should not make the assumption that we really understand what's causing these problems.

DR. GROSS: Any other comments? If not, we'll draw number 3 to a close. Okay, Brian. Robyn, do you want to go first?

MS. SHAPIRO: I just want to say that I agree and that the first thing I said this morning I feel no better about at the end of the day, and that is, that we're accepting an assumption about cause and effect that I don't feel comfortable that we can prove. Until we have our arms around that better, I don't think we could possibly answer, for example, question 5.

DR. GROSS: Well, you're going to get the last word and create a new question that we'll have to answer.

Brian.

DR. STROM: Three comments. One is as one additional thing I think we should do and which I think very much follows up on the comments that have just been made is the root cause analyses of the drugs that got into trouble with names even after the current process is over, as was suggested before.

Second is a caveat. There's been a lot of discussion about computerized order entry as the solution. We actually have data we haven't published yet of enormous numbers of errors introduced by computer order entry. So it is very far from a panacea. It solves the handwriting problem, but it introduces many, many other kind of

problems. So people should just be careful.

Third -- and this is in some ways is the opposite of Ruth's suggestion, which was obviously very well thought out and thought through, and where this is sort of seat of the pants, but it never stopped me from talking anyway. I wonder if you could take advantage -- this is not screening before marketing but after marketing, perhaps as part of risk management programs, perhaps just from a validation point of view -- using databases. For example, Avandia/Coumadin. One of the key questions that we've been struggling with today is how common are these problems. How much of a problem are they really? How many times do we see diabetics who get a single prescription of Coumadin in a database on the market or using claims data? Or how often do you have somebody who doesn't have diabetes, who is on no other diabetes drugs, who's on longstanding Coumadin, who gets a single prescription for Avandia? Those kinds of analyses would be easy to do and, in selected situations like that, could be used as a gold standard to try validate the kind of things that we've been talking about. It wouldn't work in many situations, but it would work in one like that.

DR. GROSS: Thank you all very much. We're through the first three questions. We'll reconvene at 3:15 to do the last two questions, plus a question yet-created

by Robyn.

(Recess.)

DR. GROSS: Thank you all. We're a few minutes late in getting started. The weather is approaching, so why don't we reconvene and let's begin with question 4.

I will read question 4 to you. Under what circumstances should a field test in a simulated prescribing environment be recommended? Is any one method alone sufficient as a screening tool, or should a combination of methods routinely be employed, such as behavioral testing and orthographic/phonographic testing?

We actually discussed much of this question previously. Does anybody have any additional comments that they want to make on this? Brian. I never would have guessed.

(Laughter.)

DR. STROM: I just want to go one step further and agree with what Mike was saying that I think the field test is an enormously useful idea but should not be required yet and should not be uniform. I think it needs to be evaluated and tested. To me I think it is probably the gold standard that should be used in evaluating the others and ultimately will be too impractical and too expensive to be used uniformly. So the answer to the question of under what situation should a field test be

done, I would say as part of validation efforts.

DR. GROSS: Thank you.

Eric.

DR. HOLMBOE: The only other thing I would add is I know that the FDA is currently doing something along those lines. It's listed as number 3.

I had some concerns about that just because of the numbers of people involved, the fact that there may be a bias there to begin with because you're intra-agency. So if you're going to continue that, I'd just really encourage you to look at that very carefully given you have a small n, and it gets back to Dr. Lambert's point that if you have a low frequency of events for certain drugs and you're dealing with only a small number of physicians participating, you might get into trouble.

DR. GROSS: Anybody else have any comments?

DR. MORRIS: Yes, just definitional. When I think of a field test, I think of a very, very big sample, but if you mean a simulated environment, that's not a -- as long as that's not ruled out, small samples of 50 or 100 pharmacists or doctors is reasonable and I think gives some sense of data, not just qualitative information. I would encourage that, but I agree, if we get into large amounts of money, then we're not there yet.

DR. GROSS: So there is a definitional problem

in what a field test means for the first part of the question.

For the second part of the question, from the earlier discussion I sense that the committee would agree a combination of methods, but it's hard for us at this point to define what should be in the combination. Is that fair enough? Okay.

Number 5. Yes, Lou.

DR. MORRIS: I'm pretty comfortable even at this point in saying that some combination of methods is going to be necessary. The idea that any single method is sufficient, given that we don't even know what the problem is -- I'm pretty comfortable that we're going to need a multi-factorial approach.

DR. GROSS: Yes, I think that's certainly the sense of the committee. Does anybody disagree with that?

(No response.)

DR. GROSS: Okay, fine.

Number 5. Describe the circumstances, if any, when it would be appropriate to approve a proprietary drug name. And I'll add for clarification that may cause some confusion, but it should be added "with a risk management program." Is that paraphrasing it right, Paul?

DR. SELIGMAN: Yes.

DR. GROSS: Comments? Arthur.

DR. LEVIN: When would that occur? Only if there was a breakthrough drug or something like that with the company refusing to -- I mean, you guys have the last word. Right? I'm just trying to sort of figure out when would that happen.

MR. PHILLIPS: There have been occasions where we reached an approval stage. Let's just say that we get to the final minute of an approval and we realize that we observe something now that we didn't think about. So we don't want to hold up the approval. We're not 100 percent sure that this error is going to occur. We have some doubts and the sponsor is willing to undergo a risk management program to address that concern, whatever that is. It is definitely associated with the name. So it may be that you have to do some extensive monitoring. It may have to do with setting up a surveillance system, educational campaigns, et cetera, anything that is a component of a risk management plan.

DR. GROSS: But wouldn't this be a place where you might want to do field testing to decide whether or not this was going to be an issue or not and then make a decision?

MR. PHILLIPS: Well, we would have put it through our analysis at FDA and maybe, one, there may be a difference of opinion internally at the FDA that might say

yes, we see your point, but we want to go ahead and issue the approval with a risk management plan. So maybe DMETS had a recommendation. The office, on the final approval, decides to go ahead and let it go with a risk management plan. So FDA has agreed to do this.

DR. GROSS: So it would be a post-approval --

MR. PHILLIPS: It's a pre-marketing agreement to institute a risk management plan post-marketing.

DR. GROSS: Curt.

DR. FURBERG: But how do we know that that risk management plan will work? In order to document its value, you have to spend a lot of time figuring out. So I'm not sure this is the solution. It makes me very nervous.

The only situation I can see is if you have two approved drugs and you find out after the fact that you have a problem. Before you would remove a name or change a name, you can say, well, the option is to come up with a risk management. That's the only situation I can think of.

DR. GROSS: Eric.

DR. HOLMBOE: That's exactly what I was going to say. Just, I want to second what Dr. Furberg said.

DR. GROSS: Okay.

Michael.

DR. COHEN: Perhaps this is where the laboratory and the model pharmacy might come in where they

could actually test in a controlled environment whether or not various measures that are being suggested -- other than the monitoring. For example, we've heard about tall man letters that help to differentiate one mark from another by enhancing the unique letter characters or the background of those unique letter characters, for example. That might work. There's some evidence that it does from Dr. Grasha's studies. There are other things that could be done. Another one was pre-market advertising, "coming soon" to help educate practitioners. So we just don't know how effective they are necessarily. That's the problem, but I could see where you could have a risk management plan approved for these rare cases, but exactly what they should be I guess we don't know at this point.

DR. GROSS: Yes. Jerry described some cases. Does anybody here have some other circumstances where they think this might need to be invoked? Lou.

DR. MORRIS: I was struck this morning, Jerry, when you said you reject a third of the names and then there's another class of drugs that you feel uncomfortable about. What percent do you actually feel comfortable about?

(Laughter.)

MR. PHILLIPS: No. I wouldn't categorize it that way. Out of that third, there might be some where we

have a difference of opinion on the objections.

DR. MORRIS: Okay. So what percent is it unanimous? Let me do it that way.

MR. PHILLIPS: We still reject a third. Okay?

DR. MORRIS: Yes.

MR. PHILLIPS: And for the most part, I would say probably 90 to 95 percent of those rejections are accepted by the reviewing divisions and are relayed back to the sponsors. The sponsor still can argue with us about whether we are correct or not. So you get into a discussion with the sponsors which may at this point bring up a risk management plan as a means to manage a perceived risk.

DR. MORRIS: Okay. So you're saying of the third that you would have rejected a small percentage, they come back and propose what if we do this risk management program. So that's the circumstances.

MR. PHILLIPS: That's the circumstances behind --

DR. MORRIS: It brings you up to a comfort level that you feel that it would be safe for the drug to be in the marketplace.

Does the risk management plan you're proposing also have an evaluation component or just have an evaluation component?

MR. PHILLIPS: Oftentimes we're very interested in learning the outcomes and whether they're effective or not. So that is discussed with the sponsor.

DR. GROSS: Can you give us any examples, Jerry, where this has occurred in the past with approved drugs? Or is this a theoretical thing?

MR. PHILLIPS: It's not theoretical, but I'm not sure I feel comfortable talking about it right now.

DR. GROSS: Okay, fine. I understand.

Brian.

DR. STROM: I want to go back. I strongly agree with Curt's comment, and I think it's important we keep focused on that. The purpose of risk management plans normally is to say a drug that has real benefit on one side but it has a risk, you try to reduce the risk or increase the benefit because the risk/benefit balance is a close call and a risk management plan would improve that close call.

We're not talking about a drug here. We're talking about a drug name. There's no public health benefit in having a drug name available versus another drug name. So to me the only reason one would ever do that would be exactly as Curt said, if in fact the drug is already on the market and there are side effects from a patient point of view of removing a drug name that is

already available.

I think the situations Jerry is describing I see as something different. I see it as a situation where you don't know as an agency that you want to reject it. There's not adequate data and you've decided you're going to generate some of the data after marketing instead of before marketing in order to get the answer. If there were better methods before marketing, simulations or laboratory or otherwise, you would generate those data before marketing.

But that's different from saying you have a concern about a drug name. I don't see why in the world from a public health point of view pre-marketing you would ever allow that drug name on the market. There's no positive to counterbalance the risk.

DR. GROSS: Let me ask the committee. Can we specifically answer this question or not? Can we describe circumstances in which this would occur? Jeff.

MR. BLOOM: I seem to recall in reading the review materials -- and I would certainly agree with it -- that the one circumstance that I could see where it could occur if there is a breakthrough drug that is meeting an unmet need where there is not any existing therapy for a serious or life-threatening condition. That's the only circumstance that comes to mind.

DR. STROM: But you change the name. You could still have the drug available.

MR. BLOOM: Yes. Absolutely. I agree with that, but I wouldn't want it to be held up because of a drug name, of course.

DR. GROSS: Michael.

DR. COHEN: I just have to say I think it's not so easy to say just change the name. There's a lot that goes behind that. We've heard that today too. And it might delay the drug by three months or six months or maybe even longer for all we know. I don't know everything the trademark attorneys know, but I'm sure they might run into situations like that. So I could see a public health benefit of an occasional use, a rare use of a risk management program.

DR. GROSS: Jerry, do you want to help us out on this? Give us some examples of circumstances.

MR. PHILLIPS: I'm going to give you another example. I'm not going to name the drug product, but the circumstance was a similarity with a trademark in which the product was no longer marketed in the United States, but was widely available in reference textbooks and in the literature. So within the practice setting, there was a wide recognition of this name, although it wasn't available. So there was an argument made. The risk

management plan included going and cleaning up those reference texts. It's hard to change reference textbooks that sit on our shelves.

(Laughter.)

MR. PHILLIPS: So it's an interesting argument. This is an example of how do you weigh the risk and the benefits.

DR. GROSS: You mean you're good, but you're not God.

(Laughter.)

DR. GROSS: Ruth.

DR. DAY: As I understand it, the FDA encourages sponsors to have backup names, and if the backup names went through all of the same processes that the lead name did, then we wouldn't have to wait for 3 to 6 months to switch. We'd have a backup name which was as good in many respects. Right?

DR. STROM: Plus developing a risk management plan probably wouldn't take any shorter time than testing a new name.

DR. GROSS: I'm getting the sense from the committee that it's hard to commit on this and maybe we should just say there may be circumstances in which this arises. It's hard for us to define them and if you feel you need to have a risk management plan and you have to go

through with the name and there's no possibility of changing the name at that point, then you have to do it.

MR. PHILLIPS: I think there's always the possibility of changing the name or approving the application without a name. But that presents its own problems for the sponsor for marketing the drug product.

DR. GROSS: So how does the committee want to deal with this question? How do you want to answer the question? Jackie.

DR. GARDNER: Perhaps in two parts. With respect to a post-approval situation that's been described here, I think that, as Brian defined it and Curt, if you're in a post-marketing situation, then we clearly could see a pause, a hiatus, while a risk management program is being developed before firm action is taken and, as Michael said, evaluate alternatives for the risk management program.

So in an after-market situation, a post-marketing situation, I think there are many circumstances in which it would be appropriate. Pre-marketing I have less confidence.

DR. GROSS: Jerry, does that answer the question? Paul?

MR. PHILLIPS: Yes.

DR. SELIGMAN: Yes.

DR. GROSS: As best we can. It's tough.

Robyn, question number 6.

MS. SHAPIRO: Okay. Here's question number 6. You're not going to like it.

To develop an approach to address the risk of harm related to look-alike/sound-alike drugs, is it possible -- and if so, is it advisable -- so two parts -- to pursue research or acquire data that will more precisely identify causative factors in such harm? That's my question.

VOICES: Yes.

MS. SHAPIRO: Then why aren't we talking about doing that before we get to all these other questions?

DR. STROM: We are.

MS. SHAPIRO: Did that whole proposal include collecting that kind of data?

DR. STROM: Yes.

MS. SHAPIRO: Wonderful, great. I'm happy now.

DR. GROSS: Lou.

DR. MORRIS: I disagree. I think what you were talking about, Brian, was validation processes.

MS. SHAPIRO: That's what I thought.

DR. MORRIS: And what Robyn is saying is causative factors for medication errors per se at a much more specific level, and I'm with her. I think that that's another research agenda that we should recommend.

MS. SHAPIRO: I don't know, although Curt is helping me along with my thinking here, how you can do any of this without doing that.

DR. GROSS: Ruth.

DR. DAY: Michael Cohen gave us an example, Robyn, which I think might help out, and that is that there were cases where there were two drug names on the market and there were a lot of errors being tracked. One drug name was withdrawn and a new name was given and there were no longer those kinds of errors.

MS. SHAPIRO: That's an example. That's great.

DR. DAY: It's not the whole answer. It's a tiny part of it, but it can't be overlooked.

MS. SHAPIRO: That's why my question acknowledges that closely related names or names that sound alike are related to harm. I think that we can assume that. It's a factor. But if we want to do a risk management approach --

DR. GROSS: I thought that was your question, what you're assuming.

MS. SHAPIRO: No. Part of the question is to develop an approach to address the risk of harm related to look-alike/sound-alike drugs. The assumption is that there is some. Is it possible, and if so, advisable, to pursue research that will more precisely identify causative

factors in such harm, that is, in harm that is related to look-alike/sound-alike drugs? So the assumption is that there's some and the desire is to drill deeper to find out, well, does that vary depending on whether we're looking at handwritten as opposed to verbal, does that vary depending on whether we have vast differences in dosages or administration routes. Let's get more precise in the factors involved so that we can be better in the risk management approach.

DR. GROSS: Yes, I think some of that has been done and a lot is still in progress.

MS. SHAPIRO: Good.

DR. GROSS: Arthur?

DR. LEVIN: It seems to me that the presentations we had on labs offer an opportunity to get at that because in a controlled situation, you can vary the variables and get a better understanding of the things you're asking about probably more quickly and less expensively than sort of going out and doing RFAs. I don't know. It might be a chance to have a down and dirty opportunity to get a little better handle on how all the variables play out in this.

MS. SHAPIRO: In a pharmacy, but I've seen a lot of errors that don't happen in a pharmacy that are terrible.

DR. GROSS: Brian.

DR. STROM: Yes. You're broadening the question to medication errors in general which clearly is appropriate and needs to be done, but realize it's a whole other field. The focus of today was on the name because that's what FDA regulates. But ARC, for example, has a close to $60 million a year budget studying patient safety issues. A substantial amount of that focuses on medication errors, and there's a lot of research underway. For example, at one of the centers for patient safety, we have studies underway looking at sleep issues, looking at things that determine, in an in-hospital setting looking at patients making errors from an adherence point of view. There's lots and lots of low-hanging fruit about why is it that there are medication errors. It's very clear that name confusion is a small part of it.

MS. SHAPIRO: But I think that I'm looking at a subset of that universe, and that is, if we take only the subset of look-alike/sound-alike, are there other factors? Again, if our task is to have a risk management approach that makes sense or, even before that, to determine whether we need one, then take that subset and look at other things so that we can be more sophisticated in making recommendations.

DR. GROSS: Louis.

DR. MORRIS: Again, I'm with Robyn. Just take a cognitive psychology look at this. Is it a pattern recognition problem, a pharmacist not looking long enough and hard enough, and if they did, would they then see it? Or is it not just the way the letters are formed, but is it some other aspect of the way they search their memory? There are lots of very specific issues that could help us understand the problem better. I asked Mike before. There are lots of problems here. We don't know that we know them all, and if we did know them, we don't know how much they contribute. So I think if we just stepped back and said, okay, what is the specific problem and understood that better, I'd be a lot more comfortable.

DR. GROSS: I think these comments are very important. I think they're a little bit beyond the scope of the questions. One of the panelists brought up to me, as far as question number 2 is concerned, how will we find out what's been decided? Can this advisory committee get a report back in three to six months as to what was decided about what study methods will be used as a minimum combination, and how will the other study methods be handled as far as proposals for future studies? What do you think, Paul? Can we get an answer? Can you just give us a follow-up in a few months as to what's going on?

DR. SELIGMAN: I'm happy to give you a follow-

up.

The challenge for us always is how to develop good practice in the context of an evolving science where there are people who are being injured or harmed and the degree to which we can foster best practice as we are developing the best science. This, of course, is the challenge to us. We're certainly happy to do our best to look at the data that are out there. We've done that in large measure already. The challenge that we face is, at least at this point in time, how to create practices -- we think internally within our own organization, we are doing I think the best practice we can in involving experts, using computational software, engaging in simulations to try to best understand where problems might occur with names, drawing on the best that's available within the current literature.

As I indicated, our ultimate goal is to try to, to the degree we can, level the playing field and ensure that industry is taking these approaches and looking at trade names beyond just their commercial value and trade name, but also to incorporate principles of safety and consideration of safety in those processes. At the end of the day, can we create a guidance based on what we know about the data to date in a way that will at least foster and improve the way all sponsors look at names that they

submit to us at the agency for review and create processes that allow some consistency so that sponsors will know the basis for which we make decisions about either accepting or rejecting such names.

DR. GROSS: Are there any other issues you wanted us to deal with today?

DR. SELIGMAN: Not that I'm aware of, no.

DR. GROSS: Brian.

DR. STROM: Just in comment to one of the things you're saying, Paul. I'm interested in the rest of the committee's comments on this, but my sense is it's premature to issue a guidance because we don't know what the best practices are is what I was hearing. I don't know if other people feel the same, or maybe I'm misunderstanding what a guidance is.

DR. GROSS: Yes. I guess, as happens to much in medicine where there aren't randomized controlled trials, decisions still have to be made. My sense is that's the position that they're in. Given what we know now, what are the recommendations they can make.

DR. STROM: Absolutely, but that's different from putting it in a guidance which I would think should be data-based. That's what I'm saying. I'm not saying you should change. I think doing what you're doing is on target. The new advances you're incorporating, I think all

that makes enormous sense. I think that's different from codifying it absent the data to know it's the correct thing.

DR. GROSS: Michael.

DR. COHEN: Peter, we spoke before about having FDA get together with PhRMA and other stakeholders. Could we set something now or at least set an expectation that that take place within the next 3 to 6 months and that there be a report back to this committee by perhaps the next 9 to 12 months at least?

DR. GROSS: I thought Paul said that he would do that.

DR. SELIGMAN: I guess the question is what's the nature of the feedback that you're looking for in this report. What are the questions that you're asking us to answer in getting together with PhRMA and other stakeholders? What are your expectations in terms of what we can produce in the next 6 months?

DR. GROSS: Arthur.

DR. LEVIN: I would agree with Paul's confusion about expectation because we've said get together, but we've also said get together so that you can start planning out the research agenda to move this along to a place where we feel is evidence with which to go out with a guidance. And that's going to take longer. I mean, just to know that

in 6 months you're going to get together with the stakeholders, great, but it's not going to move this much further. It's going to be more time. By saying this, we're delaying the process, and that's just the reality. We're not going to get a quick fix on this. The evidence base does not yet exist to make us comfortable to set up standards or criteria to form a guidance to give to industry to say this is what we'd like you to follow, and if you follow this, you'll be okay. We're not there yet and it's going to take not 3 to 6 months, but probably at least 12 to 24 months to get there.

DR. GROSS: Yes. My suspicion is to have an adequate evidence base to make recommendations on where each recommendation is based on good, solid scientific evidence, it will take a few years. In the meantime, drugs are still being approved. So some decision has to be made as to what methods will be used to clear those drugs to avoid confusion with other drugs. Again, we're in that scientific limbo where we don't have the evidence to make the kind of decisions we want to make but yet decisions have to be made.

DR. SELIGMAN: I also struggle a little bit with the kind of evidence that we would be looking for at the end of the day and would be actually interested in hearing from members of the panel as to what evidence we

might be looking for.

DR. MORRIS: But that's the purpose of this process we're suggesting. Eventually FDA is going to call for evidence in support of drug names, but we're saying we don't know what that evidence should look like. So as the first step in the process, because we don't know which of these methodologies or any other methodology might be the best evidence or combination of evidence, why not start a public process with PhRMA to decide, based on validation, what that evidence should be? What we're asking for is, rather than it just being a consensus process, that there actually be science underlying the type of evidence that you will eventually get and you go through this process of learning about what's the most valid methods before you ask for them.

DR. SELIGMAN: But I would argue that there's science, for instance, behind the computational searches, that these are indeed well-validated methods. Ultimately at the end of the day, somebody is going to have to look at that ranking of things that either look alike or sound alike and make some decisions based on input from expert panels which again I think can be constructed in a way that are well defined even though there are I think some significant issues regarding the validation of those. Similarly, one can go through a process, as we do, of

written and verbal Rxs and define that process very carefully.

But I guess we can do a lot of what I would call sort of internal validation of these techniques. The problem for us is how to externally validate them, to know that information that is generated out of each one of these components or the ultimate risk assessment, indeed, does have its intended impact of essentially preventing a name confusion.

DR. GROSS: Curt.

DR. FURBERG: My sense is that we have three silos. We have the FDA addressing the problem. We have the industry and then academia, and there's very poor communication between the three groups. Even within a silo there's a problem. You just heard about the pharmaceutical industry, that some companies are doing a lot and others are doing probably very little.

So I think what we need to do is to set up a situation. We can have a dialogue about what is being done right now and what are the lessons learned, what is working and what is not working. So focus on two things: one, on the knowledge we have and even take advantage of the FDA database, the 100 cases disapproved. We can learn from it. What are the patterns in that that we can learn from.

So that's what I think a meeting could do,

bring in the parties, have a good discussion about what we know and what we have learned, some further analyses, and then in addition, talk about the process. I'm not sure the process is well defined. You get names submitted to you and you review them, but maybe there should be something happening earlier than that. Maybe they should come to you and talk about this is how we're going to go about evaluating name confusion, and you need to have some guidance to them, what is it that they should do, what would speed up the process and make it more acceptable to you.

This lack of communication I find a little bit troubling, and that's why I suggested just get people together in a room and let them talk and you'll come up with something. Based on that, you may be able to, on existing evidence, come up with guidelines that could be refined, and I'm sure there will be areas or gaps. The other outcome would be even to learn what are the gaps and see what is essential that we focus on in the future.

DR. GROSS: Brian.

DR. STROM: I still think the conversation is necessary but not sufficient and you're not going to be able to put people in the room together and have them come up with a scientifically reasonable decision because there's no data underlying it. We've had two of those

meetings. We've proven that.

I think, Paul, you talked about there's science underlying the computerization. I think that's a perfect model. That's analogous to there's science, physiology, and preclinical data underlying why a drug might work and be safe, but yet we test it in people to find out and drugs don't survive their testing in people. The science that exists now is process-based science. What isn't there is outcomes-based science. There are lots of different ways you could generate it ranging from looking at drug names that failed in the past, looking at drug names that Jerry has rejected that industry has passed, doing some of the mock pharmacy or the laboratory kind of approaches.

We need outcomes-based data to validate what works and what doesn't work because the chances are there's a significant amount of what's being done now, which is fine, and there's a significant amount of what's being done now which is wasted effort. Get rid of the wasted effort. Require the stuff that's fine and add other things that are useful.

But you're not going to be able to know any of that without looking at gold standards -- or at least silver standards. There are no gold standards in the field, but at least silver standards as opposed to the fool's gold as the gold standard. We need to test all of

the methods that are now being used against some at least silver standard or group of silver standards, given none of them are gold standards. Until that's done, how can you codify requirements for what should be best practices? We don't know what the best practices are.

DR. FURBERG: Yes, but Brian, I don't think you can make progress by having another session or two of show and tell.

DR. STROM: I strongly agree.

DR. FURBERG: You need to get people together and define the issues and, maybe with you as one of the moderators, make sure that they stay on track and address the real issues.

DR. STROM: I agree, but the issue of that getting together isn't what's the best way to do it because then we're just going to have another show and tell. The purpose of the getting together is what is the research that needs to be done and who's going to come up with the money and who's going to fund it and what's the process and ideally come up with a joint process that everyone will be comfortable with which will validate or not the approaches that --

DR. FURBERG: But I see that as step two, sort of the future, what do you do. Right now, let's see what we have.

DR. GROSS: I'm hearing two different things. I'm hearing the science is insufficient to make a recommendation, and I think everybody seems to agree with that. But what's the corollary? The corollary status quo or what does the group think?

DR. STROM: It's status quo until we generate more science, and the priority should be in generating more science.

MS. SHAPIRO: Outcomes-based.

DR. STROM: Outcomes-based, yes.

DR. GROSS: Lou.

DR. MORRIS: There's another thing that we can recommend and that is that rather than being specific on what to request from the industry, that FDA, as part of this process, ask for some evidence from the industry at their choosing and that part of this time that we're spending validating, FDA can also be spending the time kind of internally validating industry evidence, and that there should be some requirement for some form of evidence. But what form it should be ultimately again is like a year-and-a-half out before we put a final guidance, but there will be this evaluation period for gathering new data and evaluating existing data that industry is already gathering but not submitting.

DR. LEVIN: A couple of things. One is I'm

comforted by comments from Michael and others that things are much better today than they were. By talking status quo, it's not the worst possible scenario. This is an issue. It's an issue people are concerned with, an issue people are working on, and there's a lot of room to grow. But things are being done.

I just want to sort of do a mea culpa from the IOM Committee perspective, that when we set a goal of error reduction and we tried to put some meat on the bones of the To Err is Human report, we thought it was incumbent on us to pick some concrete steps that could be taken right away. I guess perhaps we were delusional in thinking that this was a simple step, that we could suggest that it could happen right away, which was to get rid of this issue of sound-alike and look-alike drug names. Clearly, it is a complex issue and not so easy to resolve. So I want to sort of take partial responsibility for pushing this issue forward in a way that I think did not fully anticipate the difficulties in even something this well-focused.

I would again like to urge a reexamination of where things went wrong with this process, in other words, taking a look at where everything passed through the screen and got out there and all hell broke loose, and what was everybody thinking, both PhRMA and FDA, and maybe learning from the mistakes and using that as sort of a down and

dirty way to get much more focus on where we need to be looking.

The second thing I'd like to urge is the lab approaches, again, being able to, I would suggest, produce some very quick notions about a lot of things, including your concerns about what are all these factors that contribute, and if we can't weight them, how do we know how to react to the problem.

DR. STROM: Peter, I had two related comments.

One is to clarify a point I made. When I say status quo, I don't mean freeze in place. What's very clear is FDA is doing a lot of neat stuff, and by status quo, I mean keep doing that neat stuff and keep advancing the science as you're doing and the public health will improve accordingly. But don't put into codification something until we know what's correct or not.

I think using lab approaches makes enormous sense in validation, and I guess one of the things we didn't talk about before, in talking about prioritizing high-risk/low-risk drugs, is I would go to the high-risk drugs to be the drugs that you use those lab approaches in as part of those validation tests.

DR. GROSS: Michael Cohen, any comments?

DR. COHEN: No.

DR. GROSS: Stephanie.

DR. CRAWFORD: Thank you. Again, I'm piggy-backing on Dr. Strom's comments. I applaud the efforts that the FDA has done. I think the multi-faceted approach is certainly a phenomenal step in the correct direction.

As I assimilate some of the comments that were made by the speakers earlier this morning, something that came up on more than one occasion was concern about was the lack of transparency. So I think perhaps the agency needs to better articulate to the audiences exactly how it is determined which of the programs is used, what goes into evaluating, exactly what processes are used because I think that adds to the discomfort when it's not there, and also perhaps some people think it's not comprehensive enough in looking at all the alternatives. But otherwise I think these steps are in the right direction.

DR. GROSS: Does anybody have any other comments?

(No response.)

DR. GROSS: If not, then the meeting is adjourned.

(Whereupon, at 3:58 p.m., the committee was adjourned.)