FOOD AND DRUG ADMINISTRATION

ADVISORY COMMITTEE FOR PHARMACEUTICAL SCIENCE

8:29 a.m.

Thursday, March 13, 2003

Conference Room
5630 Fishers Lane
Food and Drug Administration
Rockville, Maryland 20857

ATTENDEES

COMMITTEE MEMBERS:

ARTHUR H. KIBBE, PH.D., Acting Chair
Chair and Professor
Department of Pharmaceutical Sciences
Nesbitt School of Pharmacy
Wilkes University
176 Franklin Avenue
Wilkes-Barre, Pennsylvania 18766

KATHLEEN REEDY, R.D.H., M.S., Executive Secretary
Advisors and Consultants Staff
Center for Drug Evaluation and Research
Food and Drug Administration (HFD-21)
5600 Fishers Lane
Rockville, Maryland 20857

JOSEPH BLOOM, PH.D.
University of Puerto Rico
School of Pharmacy
4th Floor, Office 416
P.O. Box 365067
San Juan, Puerto Rico 00935-5067

PATRICK P. DeLUCA, PH.D.
Professor, Faculty of Pharmaceutical Science
University of Kentucky, College of Pharmacy
327-H Pharmacy Building, Rose Street
Lexington, Kentucky 40536-0082

MARVIN C. MEYER, PH.D.
1700 SW 6th Avenue
Boca Raton, Florida 33486

NAIR RODRIGUEZ-HORNEDO, PH.D.
Associate Professor of Pharmaceutical Sciences
College of Pharmacy
The University of Michigan
Ann Arbor, Michigan 48109

ATTENDEES (Continued)

CONSULTANTS:

WALTER HAUCK, PH.D.
Professor of Medicine
Department of Medicine
Jefferson Medical College
Thomas Jefferson University
#1770, 132 South 10th Street
Philadelphia, Pennsylvania 19107-5244

(ROBERT) GARY HOLLENBECK, PH.D.
Associate Professor of Pharmaceutical Science
University of Maryland School of Pharmacy
20 North Pine Street
Baltimore, Maryland 21201

MICHAEL S. KORCZYNSKI, PH.D.
Senior Vice President/General Manager
Mikkor Enterprises, Inc.
P.O. Box 573
Lake Bluff, Illinois 60044

WOLFGANG SADEE, DR.RER.NAT.
Chair, Department of Pharmacology
College of Medicine and Public Health
Ohio State University
5072 Graves Hall, 333 West 10th Avenue
Columbus, Ohio 43210

CYNTHIA R.D. SELASSIE, PH.D.
Professor of Chemistry
Department of Chemistry
Pomona College
Seaver North, Room 219
645 College Avenue
Claremont, California 91711-6338

MARC SWADENER, ED.D.
2235 Dartmouth Avenue
Boulder, Colorado 80305-5207

ATTENDEES (Continued)

CONSULTANTS: (Continued)

JURGEN VENITZ, M.D., PH.D.
Department of Pharmaceutics
School of Pharmacy
Medical College of Virginia Campus
Virginia Commonwealth University
Box 980533, MCV Station
Room 340, R.B. Smith Building
410 North 12th Street
Richmond, Virginia 23298-0533

INDUSTRY REPRESENTATIVES:

LEON SHARGEL, PH.D., R.PH.
Vice President, Biopharmaceutics
Eon Labs Manufacturing, Inc.
227-15 North Conduit Avenue
Laurelton, New York 11413

EFRAIM SHEK, PH.D.
Divisional Vice President
Pharmaceutical and Analytical Research and Development
Abbott Laboratories
Dept. 04R-1, Building NCA4-4
1401 Sheridan Road
North Chicago, Illinois

GUEST SPEAKERS:

RICK GRANNEMAN, PH.D.
Divisional Vice President
Center of Clinical Assessment and
Senior Research Fellow
Global Pharmaceutical Research and Development

BO OLSSON, PH.D.
Microdrug Development AB
St. Larsvagen 42B
S-222 70 Lund
Sweden

ATTENDEES (Continued)

GUEST SPEAKERS: (Continued)

LEONARD WARTOFSKY, M.D., M.P.H., M.A.C.P.
Professor of Medicine
Chairman, Department of Medicine
Washington Hospital Center
110 Irving Street, N.W.
Washington, D.C. 20010-2975

FOOD AND DRUG ADMINISTRATION STAFF:

WALLACE ADAMS, PH.D.
BARBARA DAVIT, PH.D.
DALE CONNER, PHARM.D.
AJAZ S. HUSSAIN, PH.D.
STEVEN JOHNSON, PH.D.
LARRY LESKO, PH.D.
NAKISSA SADRIEH, PH.D.
DONALD SCHUIRMANN
SOLOMON SOBEL, M.D.
YI TSONG, PH.D.
HELEN N. WINKLE

ALSO PRESENT:

WILLIAM H. BARR, PHARM.D., PH.D.
Professor and Executive Director
Center for Drug Studies
School of Pharmacy
Virginia Commonwealth University

SANFORD BOLTON, PH.D.
Visiting Professor
University of Arizona

ROSALIND BROWN, M.D.
Children's Hospital, Boston

RICHARD DICKEY, M.D.
The Endocrine Society

CARLOS R. HAMILTON, M.D., F.A.C.E.
American Association of Clinical Endocrinologists

ATTENDEES (Continued)

ALSO PRESENT: (Continued)

BRYAN HAUGEN, M.D.
School of Medicine
University of Colorado

JAMES HENNESSEY, M.D.
Associate Professor of Medicine
Division of Endocrinology
Brown Medical School

IRWIN KLEIN, M.D.
Professor of Medicine and Cell Biology
NYU School of Medicine
Chief, Division of Endocrinology
North Shore University Hospital
Manhasset, New York
Consultant, King Pharmaceuticals, Inc.

JACOB ROBBINS, M.D.
Scientist Emeritus
National Institutes of Health
American Thyroid Association

OMEGA LOGAN SILVA, M.D.
Professor Emeritus of Medicine
George Washington University
Past President, American Medical Women's Association

R. MICHAEL TUTTLE, M.D.
Associate Professor of Medicine
Cornell University
Memorial Sloan Kettering Cancer Center

LAWRENCE WOOD, M.D.
CEO and Medical Director
Thyroid Foundation of America
President, Thyroid Federation International
Clinical Practitioner, Thyroid Unit
Massachusetts General Hospital

C O N T E N T S

P R O C E E D I N G S

(8:29 a.m.)

DR. KIBBE: Ladies and gentlemen, I want to welcome you to the second day of the meeting.

If the members of the committee will make sure they're in position and we'll get started. We have an extremely busy day. We have lots of presenters during the open discussion. So we need to be efficient, if at all possible.

Ms. Reedy will read a statement on conflict of interest.

MS. REEDY: Acknowledgement related to general matters waivers, Advisory Committee for Pharmaceutical Science on March 13th, 2003, the open session.

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 topics of this meeting are issues 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 topics at hand. Because they have reported interest in pharmaceutical companies, the Food and Drug Administration has granted general matters waivers to the following SGEs which permits them to participate in these discussions: Dr. Joseph Bloom, Dr. Patrick DeLuca, Dr. Walter Hauck, Dr. Gary Hollenbeck, Dr. Meryl Karol, Dr. Arthur Kibbe, Dr. Michael Korczynski, Dr. Marvin Meyer, Dr. Nair Rodriguez-Hornedo, Dr. Wolfgang Sadee, Dr. Jurgen Venitz.

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.

In addition, Drs. Cynthia Selassie and Marc Swadener do not require general matters waivers because they do not have any personal or imputed financial interests in any pharmaceutical firms.

Because general topics impact so many institutions, it is not prudent to recite all potential conflicts of interest as they apply to each member and consultant. FDA acknowledges that there may be potential conflicts of interest, but because of the general nature of the discussion before the committee, these potential conflicts are mitigated.

With respect to FDA's invited guests, Dr. Leonard Wartofsky reports that he has a consulting contract

with Abbott Laboratories. Dr. Bo Olsson reports that he is employed full-time by AstraZeneca Pharmaceuticals in Sweden, and Dr. Rick Granneman reports he is employed full-time as Vice President, Center of Clinical Assessment, by Abbott Laboratories.

We would also like to disclose that Dr. Leon Shargel and Dr. Efraim Shek are participating in this meeting as acting industry representatives, acting on behalf of regulated industry. Dr. Shargel reports he is employed full-time by Eon Laboratories as Vice President, Biopharmaceutics. Dr. Shek reports holding stock in Abbott Laboratories and Cephalon, Incorporated, and is employed full-time as Divisional Vice President for Abbott Laboratories.

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 all current or previous financial involvement with any firm whose product they may wish to comment upon.

DR. KIBBE: Thank you.

As is custom, we will ask the members sitting

around the table to introduce themselves. Before we get started on that, we've gotten a couple of pieces of paper put out for the members to look at. One is a listing of the members and their expertise, and we would like you to correct that and turn it back in before you leave, if there are corrections, and a list of acronyms for your use. It's only 82 pages long, so you know the alphabet soup in Washington, D.C. has not gone away.

Let's start with Ajaz and go around the table and introduce. Yes, I know, Helen gets to talk first, but you get to introduce first.

DR. HUSSAIN: Ajaz Hussain, Deputy Director, Office of Pharmaceutical Science.

MS. WINKLE: Helen Winkle, Acting Director, Office of Pharmaceutical Science.

DR. VENITZ: Jurgen Venitz, Virginia Commonwealth University, representing the Clinical Pharmacology Subcommittee.

DR. SADEE: Wolfgang Sadee, Ohio State University.

DR. RODRIGUEZ-HORNEDO: Nair Rodriguez-Hornedo, University of Michigan.

DR. SWADENER: Marc Swadener, Emeritus from the University of Colorado in Boulder.

DR. MEYER: Marvin Meyer, Emeritus Professor,

University of Tennessee.

DR. KORCZYNSKI: Michael Korczynski, Consultant, Mikkor Enterprises.

DR. BLOOM: Joseph Bloom, University of Puerto Rico.

DR. SELASSIE: Cynthia Selassie, Pomona College.

DR. HOLLENBECK: Gary Hollenbeck, University of Maryland.

DR. DeLUCA: Pat DeLuca, University of Kentucky.

DR. SHARGEL: Leon Shargel, Eon Labs, Inc.

DR. SHEK: Efraim Shek, Abbott Laboratories.

DR. HAUCK: Walter Hauck. I'm Professor and Head of Biostatistics at Thomas Jefferson University.

DR. KIBBE: Thank you, and I'm Art Kibbe, and I work at Wilkes University, Chairman of the Pharmaceutical Sciences Department and acting Chair of this committee.

Our first speaker will be the acting Chair of the Division, Helen Winkle, who's been acting for three years.

MS. WINKLE: Good morning, everyone.

I'm going to talk just briefly this morning about the GMP initiative for the 21st Century. As I said yesterday, I think that it's important for the committee to

have an idea about this initiative because it is such an important part of what we're doing in the center. I want to start off by saying that although the title of it has been in the press and when we started this initiative back in August of 2002, it was titled the Pharmaceutical cGMPs for the 21st Century, we actually look at it as the drug product quality initiative because, as I was mentioning yesterday, this covers far more than just the cGMPs. It covers the review aspect of quality as well. So it's basically a continuum from the day the products come in and how we look at the quality to come in for review for marketing to the day they basically are no longer on the market. So it is a continuum and we like to think of it in those terms.

I'm going to talk about the initiative. I'm going to run quickly through the various aspects of the initiative just so you'll have an idea of what it entails, and then Ajaz is going to sort of make the connection between many of the things we're going to be doing here at the advisory committee as well as on the various subcommittees.

First of all, just let me talk briefly about the goals of the initiative. It's basically conceived of to incorporate concepts of risk management and quality systems in what we do in our daily activities in the

agency.

It also includes the latest scientific advances in manufacturing and technology. We often find, as Ajaz talked about yesterday, that we sometimes feel like the industry doesn't move forward in these areas because FDA is sort of standing in their way. As Ajaz says, we don't want to be responsible for that. We're really trying to encourage scientific advances. So this is part of what we've built into the initiative.

We want to better integrate the review program with the inspection program which I've already mentioned. It's a continuum across.

We want to ensure consistency in standards. It's a very important part of how we do business and how industry and others do business.

And we want to encourage again innovation and focus resources effectively to address the most significant health risks that are out there.

Just to give you an overview of the initiative so you know what it entails, it basically applies to pharmaceuticals, biological human drugs, and veterinary drugs, and the focus is on the review, as I've already said, of drug product applications and the inspection of manufacturing facilities. The initiative is being coordinated through a steering committee which consists of

members from our Office of Regulatory Affairs, our Center for Biologics Evaluation and Research, our Center for Veterinary Medicine, from CDER, the Center for Drug Evaluation and Research, from our Office of the Commissioner with input both from CDRH, which is our Center for Devices and Radiological Health, and CFSAN, which is our Center for Food Safety and Applied Nutrition. So basically everyone in the agency is involved in this initiative in one way or another.

We really, when we started this initiative back in August, envisioned that it would take two years to really -- and I won't say finalize the initiative but to put the major part of the work into the initiative. Obviously it's something that will go on for a number of years out to really incorporate all those aspects of the initiative that are really important to ensure that we focus on the right things as far as quality is concerned.

We did provide our first six-month report in February, on February 20th, and we have done a lot of work in the six months within the agency, looking at how to make a number of changes, and I'll talk about that more.

I just wanted to quote Dr. McClellan here because I think his quotes are very significant when we think about this initiative and where it's going. He specified in his report on this in February that "using

state-of-the-art approaches to our review and inspection process means getting important new medications to patients faster." So there's more to this than just the obvious of what the initiative says. This is basically to help improve the whole area of medicine and to help the consumer.

Another one of his quotes on that day was, "FDA will focus our attention and resources on the areas of greatest risk with the goal of maximizing public health protection without impeding innovation."

Here, I have a chart which I know will be hard to read for you all. The advisory committee does have it in their handout. This is the chart of task groups within the initiative. As you can see, the steering committee oversees the activities of the various task groups. There are 14 of them on this chart. There are actually some other subgroups of these, but I'm going to go quickly through the main task groups again so you will have an idea of what we're doing under this initiative. Every one of you on the advisory committee did get a reference to the website which has the background materials for these working groups in there, what we announced on the 20th of February.

The first one I'm just going to touch basically on is the contracts management. This group was set up to

expedite external studies of key issues that need to be addressed under the initiative. Basically we're looking at two areas now, and we feel like we need help in the agency to really focus on these areas and that's why we're looking at having them done on contract. We're looking at effective quality systems practices. We want to sort of go out and look at those practices outside because obviously in setting up internal quality systems, we don't have all of the expertise inside of FDA to be able to put quality systems into practice within the agency. So we're going out to look at some of those and also to get a better handle on some of the areas that we need to focus on as far as with the industry on how we handle cGMPs and other product quality methodology.

So we'll be doing some contracts on this in the near future and from those contracts, we hope to learn a lot more on how we need to proceed in this area. As the initiative moves along, too, we'll go out for other contracts to help us in the agency in gaining more knowledge.

International. When the initiative first started back in August of last year, Dr. Crawford and then later Dr. McClellan, when he came on board, wanted to be certain that we include the scope of international in our thinking as far as this was concerned. He felt like that

there's a lot of efforts that take place, especially for industry. There's a lot of confusion sometimes between what we here in FDA do and what's done internationally, and he felt like this was an important part of what we needed to look at as we instituted more quality systems internally and as we looked at how we were going to ensure quality in the future.

We felt like it was important to have harmonized approaches as we looked at drug product quality, and we're doing some of that with working with ICH specifically in the realm of technological advances. In Brussels this summer, we'll begin talking about a lot of these areas at ICH. We're looking at other forums for harmonization, and also we want to be able to benchmark with other countries' systems, and we'll be doing that a lot, too, in the future.

Part 11, just quickly. This was an area, of course, of a lot of concern to industry and we have spent a lot of effort up front in focusing on this to be able to clarify the scope of FDA's electronic recordkeeping requirements, to provide for enforcement discretion in the areas where interpretation is unclear. We withdrew the draft guidance on the 4th of February. What we hope to do, in order to get more information out to industry and others who have to implement part 11, is we hope to have a webcast

where we can go out and provide information probably sometime in June or July.

And lastly, which of course will take a little bit longer time, is we're planning to amend 21 C.F.R., part 11, the rule and the preamble. So these are things in part 11 that we're focused on now.

Dispute resolution. One of the things we've heard time and time again from industry is the need to have some type of dispute resolution process where scientific and technical questions come up when we're doing inspections, that there is a route to come into the agency to sort of clarify that science and that's not existed in the past. So we're trying to set up some type of system or forum where we can do this internally within the agency and develop consistent policies and procedures for resolving these issues in the GMP area. Basically, we're looking to be able to have a dispute resolution process between regulated industry and the FDA and also between the components of FDA because there is a lack of consistency from center to center on how we will handle some of the scientific disputes.

483 communication. There has been a lot of concern on the part of industry about how we communicate observations on our 483, which is the form that's used during the inspection process. What we're planning on

doing is honing the language to communicate deficiencies better, again to be more consistent. Right now in order to ensure that consistency, we're actually combining this particular working group, the working group that's looking at communications on 483 and through inspections in general, with the dispute resolution group. So those two groups are working together to try and ensure that industry is better informed of the observations, that the observations are grounded in good science.

Also, the warning letter process is being looked at. We're launching a program to identify any inconsistency across program areas with respect to all drug cGMP letters. It varies now from center to center whether the warning letters, when they go out to industry, are reviewed in the centers and this is what we're working towards, is consistency along that line and planning that those warning letters will be reviewed in the centers before they go out. They'll be reviewed to ensure that the science is strong science, that it's built into the warning letters.

Manufacturing science. This is a very important thing. This is part of ensuring the efficiency and quality of pharmaceutical manufacturing and associated regulatory processes. We want to facilitate, as I said earlier, the introduction of modern manufacturing

technologies and systems. We also want to, though, be able to enhance FDA's expertise into pharmaceutical engineering and technologies. We ourselves admit that we need to strengthen here some of our knowledge to be able to better understand in some cases what constitutes really good quality of product, and we'll be working on doing that as part of this initiative.

Also, I think we talked briefly yesterday, Ajaz talked briefly about the PAT initiative, the process analytical technologies, and this is part of the manufacturing science part of the GMP initiative. I think, too, this is one part that we will see continuously with this advisory committee. We'll bring a number of questions, I think, at least to the Manufacturing Subcommittee and then on to the advisory committee.

Changes without prior review. We talked about this yesterday on comparability protocols. This is to identify opportunities to allow postapproval manufacturing changes without FDA review and approval prior to implementation.

Risk management work planning. This is an area that we feel like we need to spend efforts on in the agency. We need to have a better way of ensuring systematic risk management approaches throughout. We need to implement risk-based approaches that focus both industry

and FDA's attention on the critical areas which we don't always do, either from the review or the GMP aspect, and recently, we have reorganized, at least CDER's Office of Compliance, to better focus on how we can improve our risk management.

The pharmaceutical inspectorate. Basically what we want to do in the agency, for at least pharmaceuticals, is to set up a specific cadre of inspectors in the field who can focus and have better knowledge on drugs so that when they go out, they have a better understanding of not only the manufacturing processes but of the products themselves.

We're hoping through this to enhance the agency's expertise in pharmaceutical technologies, to ensure state-of-the-art pharmaceutical science. What we'll do is, although we do have staff in our field operations now who will move into this cadre, we're looking to enhance that staff with additional staff and to continue to increase their expertise through better training, maybe even better involvement with the industry, training through the industry facilities as well, and also establish a closer working relationship between the field and the centers.

Product specialists. What we're striving to do here is develop highly trained FDA product specialists to

basically help in strengthening consistency in regulatory decisions and ensure submission reviews and that the inspections are coordinated and synergistic. Again, we will have people in the centers, in the field, who have the technical information that's really necessary to get into the more complicated areas of manufacturing and understand those as we do inspections and reviews.

Team biologics. In the Center for Biologics Evaluation and Research, they do their cGMPs a little bit differently. They have an internal team. The team biologics has been in existence for awhile, and looking at that, how team biologics works and the effectiveness of it has been studied for awhile, and now it's been built into this drug product quality systems initiative. And basically we're looking at improving the operations of team biologics and building on the implementation of a quality management system. And as the CDER/CBER consolidation becomes effective, obviously some things with team biologics are going to change a little bit to align them with how CDER does business. So there are some areas here, too, that we'll have to focus on under the initiative.

Quality systems. Basically, we're looking both internally to set up quality systems and externally to understand better the quality systems that exist out there in manufacturing. We hope to improve both review and

inspectional processes through implementing these quality systems approaches, and as part of this, too, we'll be looking at our regulations.

Training. Basically, this affects all the areas. Everything that I have mentioned here will have a training component to it. So this is a very important part of the overall initiative, and basically we will have to take a look at what we need for training. We'll have to do training both internally and externally, and we're in the process of beginning to develop some of these training courses and determining what we really need to be doing.

And lastly, evaluation, which is an important part of any initiative, and we feel this is extremely important to the initiative. In fact, Dr. Woodcock herself is heading up this particular working group. What we hope to be able to do is to develop appropriate metrics and a mechanism for evaluating the entire initiative, so that two years from now, three years from now, four years from now, whatever, we can go back and look at how successful we have been in instituting the changes under the initiative.

Basically, next steps is we'll have a workshop in April to begin to vet a number of these initiatives, to get input from the stakeholders. I think this is an important part of the overall initiative. We'll also be vetting a number of the questions, scientific questions

that come up in the area of manufacturing before the subcommittee and the advisory committee. As I've said, I think you'll see a number of these issues in the next six months or so.

We're getting several draft guidances out to issue for public comment, including the one on comparable protocol and dispute resolution. We'll definitely have additional workshops to focus on a number of the scientific issues under the initiative, probably even have another workshop before the year is over, and again we're in the process of clarifying part 11.

So these are just the immediate steps. Obviously, as the initiative continues to gain momentum, there will be a number of other things that will be added to this list of steps, but we've all been very active and busily working on this initiative. And again, I think it's important because I think, as I said yesterday, we're going to start seeing the scientific environment anyway of the agency change and this initiative is really an important part of those changes.

So anyway, I thank you. Again, it was a lot to listen to. There is a lot going on here. So I appreciate your attention, and I'm going to hand it over to Ajaz.

DR. KIBBE: Thank you, Helen.

Is there anybody that has any questions before

you sneak away?

(No response.)

DR. KIBBE: Your presentation must have been perfect.

DR. HUSSAIN: I'm going to continue with your advice and not use slides.

Let me start where Helen stopped. The workshop, the inaugural workshop for this initiative is on April 22nd to 24th. We anticipate this to fill up quickly. So if you haven't registered, you should register as soon as possible. The registration information is available on the FDA website as well as the PQRI website. This workshop is designed to get input from industry and other stakeholders, and we'll have a very interactive session which will be in four parts, sort of breakout sessions in four different areas. These areas are risk-based GMPs, defining risk and quality, integrated quality systems, focusing more on review inspection, and changes without prior review and manufacturing science. So if you have not registered, please do so quickly, and the number of slots available will be limited. We anticipate this to sell out.

As part of this initiative, we have defined from an FDA perspective a vision for the future, what we would like to see or what we anticipate the future to be in terms of manufacturing, and I think it's important to focus

on that and how do we get there depends on what we do today. So all the activities, discussions that we had yesterday and we'll have today impact on the future state, and what I would like to do is sort of walk through the future state that we think is a desired state and then try to link yesterday's discussion and today's discussion to that and hopefully connect those dots.

I think the drug discovery development paradigm is shifting, and one anticipated outcome is that the trend would be more towards targeted small populations and drugs developed for those, and I think that itself creates a challenge, and manufacturing would have to be flexible to adapt to that. At the same time, I think efficiency of manufacturing processes need to be at a much higher level for many different reasons.

So in the drug quality system for the 21st century, we essentially want to recognize that pharmaceutical manufacturing is evolving from an art form to one that is now science- and engineering-based. Effectively using this knowledge in regulatory decisions, not only for establishing specifications but also for evaluating manufacturing processes, can substantially improve the efficiency of both manufacturing and regulatory processes.

This initiative is designed to do just that,

through an integrated systems approach, to product quality regulation, focused on sound science and engineering principles for assessing and mitigating risk of poor product and process quality within the context of the intended use of pharmaceutical products. And with that sort of a framework, I think what is the desired state for pharmaceutical manufacturing from development and manufacturing?

One, product quality and performance achieved and assured by design of effective and efficient manufacturing processes. The emphasis there on design is to sort of emphasize that testing to document quality is not a paradigm which really is the current state of thinking. It has to be by design.

Product specifications, based on a mechanistic understanding of how formulation and process factors impact product performance, continuous real-time assurance of quality, regulatory policies tailored to recognize the level of scientific knowledge supporting product applications, process validation and process capability, risk-based regulatory scrutiny that relates to, one, level of scientific understanding of how formulation and manufacturing process factors affect product quality and performance, and two, the capability of process control strategies to prevent or mitigate risk of producing a poor

quality product.

So this is where we want to be in the future and what we have to do today and how do we get there, I think we will be seeking your input on that in that journey.

Yesterday we discussed many topics which I think you can now link this to the future state. For example, yesterday we discussed our system for ensuring therapeutic equivalence of generic drugs and also innovator drugs in the event of postapproval changes. One topic that we discussed yesterday was topical products nomenclature that dealt with pharmaceutical equivalence, bioequivalence, and therapeutic equivalence, for example.

I also pointed out yesterday that if we do not look at that from a systems perspective, there is a humongous potential for misunderstanding, and if you just focus on bioequivalence, bioequivalence is never equal to therapeutic equivalence. That's not the mantra we have. That's not our system. Our system starts with an entire assessment of pharmaceutical equivalence, manufacturing process, labeling. These are all components to that that makes a decision whether a product is therapeutically equivalent or not.

We also discussed yesterday the concept of the comparability protocol which is directly linked to this,

but at the same time, I think when you look at the information base that we use to set specifications and identify critical formulation variables and so forth, there's a lot of information that exists today that is not effectively used.

One of the concepts that was discussed yesterday was design your own SUPAC or make your own SUPAC or customized SUPAC, whatever you would like to call that. That is based on an understanding of your manufacturing process variables which are critical in how they impact on product performance. If we effectively utilize that information, I think we can do a much better job in managing changes, and why are changes important? Change is a way of life. In fact, changes are the only way forward, and when there is a change in manufacturing process or when there is a change in the product composition, I think clearly the concern from the public health perspective is that this change should not affect the safety and efficacy profile. And that is the challenge that FDA and the industry have.

I think we need to find effective and efficient methods for ensuring that product performance is unchanged and the manufacturing process changes that occur keep improving the efficiency, and that's sort of a continuous improvement model that comes about. So that's a challenge

and that's what we discussed yesterday.

Today, we'll discuss a proposal on a parametric tolerance interval approach to dose content uniformity. In fact, if you go back and recall, one of the slides Tom Layloff presented in his presentation on content uniformity for tablets, it's a direct link to that. I'm very excited about this proposal. Conceptually, I think we are in agreement that this is the direction we would like to go, and why are we so excited about this proposal?

In Tom Layloff's presentation, you saw our current approach to many of the tests that we have, say, in the USP content uniformity are zero tolerance tests. USP tests were essentially evolved as a market standard where a pharmacist or physician can take 10, 20, 30 tablets and say yes, no doubt it is outside 75 to 125.

The parametric approach that you will hear today, I think, is an evolutionary step in sort of bringing the current state of statistical science to bear on certain decisions, and you actually take into consideration the variability, the underlying distributions, and actually you can make better decisions with this.

That is so critical as we move towards the future. The reason is, if you have now the capability, say, with the process analytical technology to essentially do a test for an entire manufacturing product lot non-

destructively, the USP-type specification is not conducive to that sort of an assessment. So you really have to take the next evolutionary step and bring a sound statistically based approach to doing that assessment and you'll hear that proposal after my presentation today.

I think one of the challenges there is there are two issues being discussed with that proposal. One is moving towards the parametric tolerance interval criteria. That's wonderful. The other aspect I think where we are struggling internally is how do you establish the acceptance criteria? So if you think about and listen to that presentation, which is an awareness topic -- and you'll have a much in-depth discussion at a subsequent meeting -- think of that as two areas, moving towards the parametric tolerance interval and then establishing what are the acceptance criteria.

The other presentation we'll have today is on endogenous substances, bioavailability and bioequivalence of that, and we discussed this yesterday, also. I think many issues remain unresolved with respect to bioavailability/bioequivalence, many are perception issues, many are scientific issues. And I think the Biopharmaceutics Subcommittee will have to prioritize and start moving towards that. This could be a topic, one of the topics, for the Biopharmaceutics Subcommittee, to come

up with a general decision tree criteria of how we approach endogenous substances. Today we do that on the basis of each product, each drug, and I think we're very confident that our system works. But I think it would be helpful to move from going for each drug-specific issue to create a framework of a decision tree. So the discussion on that is focused on where do we go from here to a decision tree criteria.

We'll end this day with a look at some of the activities, research activities in our immediate office. There are two points that I would like to make with that.

One is as we move towards a quality system approach to thinking, there has to be a mechanism for evaluating how good we are. We, for several years, had a committee called Therapeutic Inequivalence Action Coordinating Committee. We talked about it briefly yesterday. What Helen has asked me to do is to take responsibility for that committee and we have taken a step back to evaluate how best do we assess and evaluate and manage that process? What is that process? It is a quality systems process, if you think of it. We get consumer complaints. We get complaints that this product didn't work as it was expected to, and how do we resolve that? How do we distinguish between whether this is a perception issue or whether it's truly a quality issue and

truly that we need to change? We took a step back, looked at the whole process, and we will sort of bring some of that discussion to the Biopharmaceutics Subcommittee, also.

But some of the research activities at the OPS level are focused on rapid response situations. This is one of the examples of the rapid response things that we do, but there are others. Some of them are related to counter-terrorism issues and Nakissa will give you some examples so that you appreciate the quality systems approach that is evolving, which is also sort of building on what we have today.

So that's what we have in store for you today, and I hope it will be a very productive discussion.

Thank you.

DR. KIBBE: Thank you, Ajaz.

Is there anyone who has any questions for Ajaz?

(No response.)

DR. KIBBE: Okay. We're scheduled to take a break at 9:30 and it is 9:08. There are a few things that we can do during that break and then perhaps we could get started with the next set of speakers a little sooner and that would give us a little more breathing room. We have, I think, 12 or so people who have scheduled to speak during the open public hearing.

Those who have scheduled to speak at the open

public hearing, if you're here and you're ready to start early, if you'll be prepared to go when we finish with our next topic, that would be greatly appreciated. Also, if you have not already checked in with staff, Kathleen Reedy would like to see you to make sure the slides and everything are all lined up.

For the members of the committee, don't forget to fill out your little lunch thing and they'll be around to pick that up.

And all the copies of everything that we have that we're looking forward to hearing today are either in your little purple folders or copied for you. If we get additional stuff, we'll get it out to you.

That being said, why don't we take a 15-minute break and come back at 9:23.

(Recess.)

DR. KIBBE: If we could start to settle down or settle down to start or whichever way you want to put it.

I have been informed that I cannot start the open hearing sooner to try to fit more time in for our speakers because of the way it is announced in the Federal Register, and so it has to start at exactly 11:30, no sooner, which means that Dr. Adams and his colleagues will have additional time to more completely describe for us dose content uniformity, parametric tolerance interval test

for aerosol products, and I think we'll benefit from that, as soon as the electronics are ready.

Dr. Adams, you're on.

DR. ADAMS: Yes. Thank you, Dr. Kibbe.

Dr. Kibbe, advisory members, good morning. I'm pleased to be here and have an opportunity to discuss the dose content uniformity work which we have been involved in for a period of time.

I'd like to note that this topic, at least my presentation, is called dose content uniformity for aerosol products, and while the approach could apply to other dosage forms as well, why aerosol products? Well, it goes back to mid-1997 when the office and the center formed an OINDP Technical Committee, Orally Inhaled and Nasal Drug Products Technical Committee, and then in 1998, a group of us within that technical committee considered batch release for dose content uniformity and whether a test could be improved. What we were looking at was dose content uniformity in the perspective of orally inhaled and nasal drug products; that is, the entire range of metered-dose inhalers, dry powder inhalers, nasal sprays, and concentrating on that effort.

Why aerosol products? It's because these products are a combination -- they're not only formulations but they're formulations with a device. So it's a drug-

device combination product, and as such, there can be greater challenges with regard to dose uniformity, both in mean delivery and in variability. So we concentrated on that effort and felt that there was an opportunity to improve the presently used dose content uniformity test.

As Dr. Hauck will indicate in his presentation, the current test specifies what constitutes an acceptable sample, but it does not indicate what constitutes an acceptable batch.

Now, there are two guidances which are appropriate to this topic. One is the Metered Dose Inhaler and Dry Powder Inhaler Drug Products-CMC documentation draft guidance issued in 1998, and then a second guidance, the Nasal Spray and Inhalation Solution, Suspension and Spray Drug Products-CMC documentation. That's a final guidance and that was published in July of 2002, and both of those guidances cover dose content uniformity recommendations.

Now, this slide is simply a nomenclature slide to indicate that the first guidance, the MDI and DPI guidance, refers to dose content uniformity and the nasal spray guidance refers to spray content uniformity. Uniformity of metered doses from an MDI, DPI or nasal spray considers performance within a container for multiple-dose products, among containers, and between batches.

The present DCU and SCU tests are essentially nonparametric tests, but they do have a parametric element. They apply to single-dose aerosol products and they apply to multiple-dose products. It's a two-tiered test as it's presented in the guidance, and at tier 1, it says that there's not more than 1 of 10 containers outside of 80 to a 120 percent of label claim and 0 outside of 75 to 125 percent of label claim. That's what we call the zero tolerance criterion, and it's an attempt to use the sample but to provide some assurance that there will not in the batch be samples with very high variability.

The parametric element in that test is the last line indicating that the mean of the 10 samples at the first tier shall not be outside of 85 to a 115 percent of label claim.

In addition to that dose content uniformity test, there's an additional test for multi-dose products and that additional test is called the Dose Content Uniformity Through Container Life for Multi-Dose Products, and for metered-dose inhalers, that test says that the dose content uniformity is measured at the beginning, middle and end life stages.

Now, for multiple-dose products, like, let's say, albuterol MDI, where the standard product is labeled for 200 doses, it's saying after priming, we want the

information in terms of dose content uniformity at the first primed dose, somewhere in the middle, and then at the 200th dose. So the goal there is to look at variability within the container. So that's why beginning, middle and end life stages is included.

The test calls for that information to be conducted on each of three containers. That's a total of nine determinations at tier 1, and similar to the prior recommendation, not more than one of the nine determinations shall lie outside of 80 to 120 percent of label claim, zero tolerance criterion, 0 outside of 75 to 125 percent of label claim, and again the means at each of the beginning, middle and end are not outside of 85 to 115.

This test simply indicates that this DCU through container life for the multi-dose products applies also in its essential characteristics to dry powder inhalers and also to nasal sprays.

Now, there have been a number of publications talking about parametric tolerance interval tests for various dosage forms, and a parametric tolerance interval approach takes the general form of the criterion indicated here that equals Y plus or minus kS, where we're defining Y, for dose content uniformity specifications, as being the absolute value of the difference between the label claim and the sample mean. And my equation really should be

slightly modified in that because I'm talking about an absolute value, it doesn't need that minus sign. It should just be Y plus kS really, if we talk about the absolute value.

K is the tolerance interval constant. The S is the sample standard deviation, and the acceptance value for this approach says that the acceptance value is less than or equal to Y plus or minus -- that is, Y plus or minus kS is less than or equal to the tolerance interval limits. I think that will be a little clearer as we proceed.

A parametric tolerance interval test, based upon hypothesis testing, is intended to control the ranges of specified coverage; that is, it may say, for instance, 85 percent of the doses in the batch fall within 75 to 125 percent of label claim at 95 percent confidence, and therefore we're specifying some minimum proportion of the batch that should fall within the limits. That's called the coverage. We're specifying the acceptable tolerance limits, the target interval -- in this case 75 to 125 percent is shown -- and the degree of confidence. That's an alpha level of 5 percent or less.

Now, a little bit of history in terms of these publications. A tolerance interval approach is official in the Japanese Pharmacopeia for a variety of dosage forms unspecified. That was based upon the work of the Japanese

statistician Katori, et al., and it is now official. It has been official since 1996. The pharmacopeia discussion group which consists of representatives of the EP, the JP, and the USP, has published on this topic. The Statistics Working Group of PhRMA has published on this topic. They have three publications in the Pharmacopeia Forum, and ICH/PDG Task Force has published and in fact has the latest article in a year 2002 issue of the Pharmacopeia Forum. All of those applications of the tolerance interval are not based upon hypothesis testing.

The first bullet here refers to a publication of Roger Williams, Guirag Poochikian, Walter Hauck and myself, published in 2002, Content Uniformity and Dose Uniformity, Current Approaches, Statistical Analyses and Presentation of an Alternative Approach, with Special Reference to Oral Inhalation and Nasal Drug Products, again with special reference to the OINDP. This paper proposed an approach that clearly states the allowable level of consumer risk and of what constitutes an acceptable batch. It didn't state what constitutes the acceptable batch, but it proposed an approach that allows for specification of an acceptable batch.

Then, lastly, on November 15th of 2001, IPAC-RS presented to the agency a lengthy report called A Permit to Tolerance Interval Test for Improved Control of Delivered

Dose Uniformity of Orally Inhaled and Nasal Drug Products, and that also is based upon hypothesis testing, and it includes, in addition to the tolerance interval, two side conditions. One is a limit on the standard deviation and another is a limit on the mean, and Dr. Olsson will discuss that in more detail.

Now, I've now got a series of four slides outlining OPS issues that has been discussed in earlier meetings between the agency and IPAC-RS, but before I present these four issues, some of which may in fact have been addressed by IPAC-RS and Dr. Olsson will talk to these issues, but before I do that, I'd like to say that OPS is interested in implementing a parametric tolerance interval approach for dose content uniformity. It places the test on a firm statistical basis and by that, I mean, it clearly states the allowable consumer risk; that is, an alpha of not more than 5 percent. It clearly specifies a limiting quality standard. It allows firms to control producer risk through selection of sample size and number of tiers of testing, and as proposed by IPAC-RS, it eliminates the zero tolerance criterion, and we know that the zero tolerance criterion represents a problem as n increases; as the sample size increases, there's more likelihood of finding a particular sample outside of that tolerance limit, and Dr. Hauck will describe that issue.

But for the above reasons that I just mentioned, we do view that should such a test be implemented, it would represent a win-win for both consumer and industry.

But I want to indicate that there are certain issues that remain to be resolved at this point, and we are simply bringing this topic to the advisory committee as an awareness issue at this time.

The first one. Dr. Hussain has spoken to this issue a few minutes ago when he indicated that the definition of limiting quality has not been resolved. There are a number of choices, based upon this parametric tolerance interval approach. One is the approach which IPAC-RS proposes. That's the first bullet. 85 percent of the doses of the batch to fall within 75 to 125 percent of label claim.

But there are other definitions of limiting quality which could be used. One is that 85 percent of the doses fall within 80 to 120 percent, a narrower range, of label claim. Another is that even more samples, 90 percent of the doses could fall within 75 to 125 percent of label claim, or 90 percent of doses might fall within 80 to 120 percent of label claim. And there may be other options for that. But that is not a settled issue and that is one of the main issues that we continue to work with on this

issue.

Another issue is robustness of the test. There are questions for non-normally distributed data and, for instance, for short-tailed distributions, and I'm aware that Dr. Olsson will be speaking to this issue of the non-normally distributed data.

Properties of the test when the batch is at or below the IPAC-proposed limiting quality of 85 percent coverage.

Another issue is the impact of eliminating the zero tolerance criterion. IPAC-RS claims that this criterion increases the producer risk with little improvement in consumer protection, but it may have some value for skewed data; that is, the distribution which is non-normally distributed and some data which are way out. So it may have some value in protecting against skewed data.

And lastly, the issue of the alpha level being less than or equal to .05 percent. We did some analyses of this approach in house. Don Schuirmann did this work and found that under certain circumstances, in fact, the alpha level goes considerably higher than 5 percent, and subsequently, IPAC-RS addressed this issue and has now reduced that alpha level closer to 5 percent, perhaps slightly above, but it all depends upon the particular non-

normal distribution and the distance between the label claim and the mean.

So what approaches are there to assuring an alpha of .05? Dr. Hauck, I believe, is also going to speak to that issue.

I'd like to finish up then with two questions to the advisory committee. We will come back to these after hearing the presentations by Dr. Olsson and Dr. Hauck. The first question for the advisory committee -- I think we'll be putting this up on the screen later -- is, does the ACPS agree that a parametric tolerance interval test is conceptually acceptable as a replacement for the agency's non-parametric DCU and DCU through container life tests for OINDPs? And to help the committee answer this question, as I say, we've asked Dr. Bo Olsson, representing IPAC-RS, to describe their approach to us.

I'd also emphasize that the IPAC-RS approach is claimed to be based upon the current FDA/DCU acceptance rule, but certainly as we'll see, the operating characteristic curves for the FDA's test and the IPAC-RS test are not superimposable.

Then following Dr. Olsson's presentation, OPS has asked Dr. Walter Hauck to provide us with his assessment of the PTIT issues and how the IPAC-RS approach deals with them.

And then question number 2 is an issue that was raised by Dr. Hussain, and it has to do with a validation of manufacturing processes issue. It says, does ACPS feel that the DCU quality standards should provide an assurance that batch failure rates do not exceed some specified level, e.g., 10 percent?

The genesis of that question comes about from a court decision back in February of 1993, Judge Wolin, who said the following, and I'm paraphrasing. The government first argues that the failure rate associated with the firm's products demonstrate the need to revise the underlying manufacturing processes. To the extent that batches included in retrospective studies exhibit a failure rate of 10 percent or more, the court agrees. So, therefore, we've been looking at this 10 percent issue and trying to determine if somehow this level of protection could be built into this test.

Now, we could look at this in a couple of ways. One is to say that the DCU test is only one of a number of tests that these products must meet in order to be acceptable. Another important one for aerosol products, in addition to the dose content uniformity, is the particle size distribution. But it seems to me that very tight specifications could be set on a DCU test and yet tell us nothing about the goodness of the particle size

distribution, and so I think they're independent tests. So how does that fit into this issue?

And secondly, if we look at this 10 percent level as applying only to the parametric tolerance interval test, is there some way that we might be able to address this 10 percent issue in setting specifications on the parametric tolerance interval test?

With that, I'd like to stop and finish up with an acknowledgement slide, acknowledging Dr. Hussain, Dr. Poochikian, Mr. Schuirmann, Dr. Meiyu Shen, Dr. Yi Tsong, all from FDA, to acknowledge Dr. Walter Hauck, who's been involved with this issue when it was first raised under a contract that the agency had with Dr. Hauck, and lastly Dr. Roger Williams, who was the individual who back in 1998 had raised this issue when he was the OPS director and was looking at approaches that may be suitable for improving the statistical basis for dose content uniformity.

Thank you.

DR. KIBBE: Do you want to take questions now or do you want to take them after your other two speakers?

DR. ADAMS: I think it might be appropriate, Dr. Kibbe, if we took them later, but it's up to the chair and it's up to Dr. Hussain.

DR. HUSSAIN: I just want to introduce the two individuals to my right. Don Schuirmann and he will

participate in the discussion of the committee this morning.

DR. KIBBE: Thank you, Ajaz.

Dr. Olsson, I think we're --

DR. ADAMS: Yes, Dr. Olsson is up next.

DR. KIBBE: Good. Thank you.

DR. OLSSON: Good morning, ladies and gentlemen, and I think I'd like to start out by thanking the FDA for this invitation to give me the opportunity to speak about the parametric tolerance interval test for improved control of delivered-dose uniformity in OINDPs.

I will only, of course, give you an overview here. You have a lot of data in the material that's in your background packages. I will try to address each of the issues that the agency and Wally in his presentation here have raised, and as he indicated, some of the answers to those issues have been recently provided to the agency in a package that I do not think that you have received yet. At the end of this presentation, I do hope that the advisory committee will agree that the PTI test is a step forward.

As we heard Wally tell you, the DDU is one of several quality attributes that is tested for OINDPs, and importantly, this one combines the performance of the delivery device and the formulation which makes it a more

complex thing, and DDU is there to verify delivered-dose uniformity in the batch, between containers and within containers for a multi-dose product, and, of course, closeness to the target.

So there are many types of oral inhalation and nasal drug products: pressurized metered-dose inhalers, dry powder inhalers, nasal sprays, inhalation solutions. All of them are intended to deliver a dose of aerosol to the respiratory tract to treat different diseases.

Ever since its introduction in the '50s, the CFC pMDI has been the main formulation type of aerosols. CFCs were linked to ozone depletion and are now being phased out. This phase-out of CFCs forces reformulation and development of new technologies for aerosol delivery.

The regulatory requirements for delivered-dose uniformity evolved mainly based on FDA's experience with these CFC pMDI products. Over time, the DDU testing requirements became more stringent. Now, even for the mature technology of CFCs, this poses challenges, and even more so with the new technologies where formulation options are more limited.

I don't think I need to go through this slide in any detail because Wally did that for me. Thank you. I just want to highlight this undesirable characteristic of a zero tolerance requirement; namely, that the stringentness

of that requirement is completely correlated to the sample size. So the more you look, the more certainty you have in failing that requirement. Therefore, it is unsuitable for situations where you do a lot of testing, for example, in stability testing, in validations, and as Ajaz pointed out, in PAT.

The reason that IPAC-RS would like to see a change of the draft guidances and the replacement with the PTIT is that because the PTIT is a more powerful test. It uses the data collected in a more efficient way and it does not have this penalty with increased testing. Another main reason is that many of the OINDPs cannot routinely meet expectations in the draft guidances, and this is demonstrated by the fact that for many products, there have been approved exceptions and deviations from the test and acceptance criteria in the published guidances.

The statistical design of the PTIT is built on previous work, mainly by Dr. Walter Hauck, but also work performed within the pharmacopoeias and especially the Japanese Pharmacopoeia, but it also incorporates some features of the FDA draft guidance test. The acceptance criteria were designed to match or exceed the statistical consumer protection implied by the published guidances.

Briefly, the batch quality definition is based on coverage, which is the proportion of doses in the batch

are within a set target interval. This means that batches having the same coverage of a given target interval are considered to be of equal quality, and this provides the simultaneous control of the closeness to the target and the variability around the mean. So when the mean drifts away from the target, then the standard deviation has to be lower in order to maintain the coverage.

Similarly to bioequivalence testing where inequivalence is the null hypothesis, we have defined null hypothesis as a batch quality out of specification. This associates the type I error with the practical most important error; namely, the undesirable event that a batch is released but is outside specification. This is yet not the usual approach within the CMC arena as it is in clinical sciences, but it is necessary to provide statistical rigor.

Since the quality of batches released to the consumer is of the greatest importance, it is appropriate to set the null hypothesis at out of specification because this then has to be refuted by data with high confidence in order for the batch to pass. And this is key to understanding our approach to the view, and I hope that Walter will touch upon this hypothesis framework a bit more, so it will be crystal clear at the end of the day.

Our proposed standard of quality is as Wally

indicated. 85 percent of batch coverage of the 75 to 125 percent label claim target interval should be covered and this corresponds to the 5 percent acceptance point for the FDA multi-dose product test. Importantly, this means that commercial batches must far exceed the 85 percent coverage; otherwise the reject rate would be unacceptably high.

So here's a comparison between the coverage at the limiting quality between the FDA and the PTI tests. So the PTI proposal is the same coverage as with the FDA test for multi-dose products and exceeds that for single-dose products.

This is a summary of the actual mechanics of the PTI test. You test a predefined number of units and those are from different portions in the container life, if it's a multi-dose product, one dose from each unit. From this sample, one calculates the mean and the standard deviation, and this is what makes this test a parametric test because these are the parameters of a normal distribution.

From these parameters, an acceptance value is calculated, and the acceptance value is the deviation of the mean value from the target, which is a 100, plus the standard deviation scale with the test coefficients.

Then the three metrics are compared with their limits, so the acceptance value needs to be lower than 25,

which is the target interval, the mean should not deviate more than 15 percent label claim, and the results are the limits on the standard deviation, which is scale with the test coefficients.

These test coefficients are listed here, and they vary with the sample size in order to ensure the type I error to be at about 5 percent at the limiting quality for all sample sizes. This means that the consumer protection is the same for all sample sizes by design but that the producer risk varies with sample size and is decreased when the sample size increases. This provides for the opportunity to select the test plan or a sample size that is appropriate for each product.

As Wally explained, these test coefficients were recently revised to address some concerns by the agency and that was to make sure that the 5 percent type I error rate was not exceeded when batch means went off the target. And here's a plot to show the acceptance probability versus the batch mean for a number of sample sizes, and this shows that only for batch means at around 9 percent deviation from the target does the type I error at the limiting quality approach or slightly exceed 5 percent. So this addresses one of the issues.

The other issues are listed here, and I will spend the remainder of my presentation going through the

bolded points here.

Just a quick note on representative sampling. This is an issue that is as important for any test whatsoever and has nothing specifically to do with the PTI test. And IPAC-RS, we do absolutely agree that representative sampling is a necessary prerequisite for any test.

Also a quick note on the topic of differences between product types, to tell you that with the PTI test where the sample size can be adjusted without compromising consumer protection, this test is well suited to take care of differences between different product types yet having a consistent standard.

We've had several meetings with the agency to discuss and resolve issues with this test. I think it's fair to say that we have reached an understanding that conceptually the PTI test is acceptable and that the main question that needs resolution is the acceptance criteria to be used with the test.

Now, let's talk about the gap which is really about the sameness or comparisons between the PTI test and the FDA draft guidance test. But first, let me go through a generic operating characteristic curve.

We have here probability to accept as the y axis and some batch variability measure along the x axis,

so that we have low variability here and high variability here. So for low variability, that is really the producer protection region, and I should say that this curve here traces the probability that the sample obtained from a batch of the corresponding batch variability is within the specified acceptance limits. So it's the ability of the batch to provide a sample within the limits that makes up this curve.

So in the producer protection region, ideally, the acceptance probability should be a 100 percent for good quality and deviations from a 100 percent. That is what we call the type II error, or beta error. As variability is increased and you come into a region with unacceptably high variability, that is where you need your consumer protection, and ideally here, the acceptance probability should be 0 and deviations from this ideal 0, that is the type I error, or alpha error.

Now, as the curve transits from the high acceptance region to the low acceptance region, there is an area of uncertainty which is where the acceptance probability is neither good nor bad. Of course, the steeper the curve, the smaller is this area of uncertainty.

This is a very important slide. This shows the comparison between the PTI test curve for a sample size of 12/36 with the draft guidance test curve for multi-dose

products. Importantly here at 5 percent acceptance rate, which is the same to say 95 percent rejection probability, the two curves tie. So they have the same consumer protection or, in other words, they have the same ability to reject quality of this type.

The PTI test is sharper. It's more discriminatory, and that is why this curve is above that of the FDA curve in the producer protection region. So fewer acceptable batches are rejected by the PTI test. This means that the producer risk is lower. The gap is due to this more efficient discriminatory power of the PTI test and it's there by design. This is what we want. The gap is not an incidental feature of the test. Industry needs to be able to approve products, if that product is of acceptable quality.

Another important point is that this curve here represents the draft guidance test curve exactly as written in the guidances. That is not to say that it necessarily reflects the OC curves of the specifications for approved products on the market.

Now, this plot here shows three theoretical examples of the effects of the types of deviations that have been approved by the agency in the last decade. We can see that the gap between the FDA curve with deviations and the PTI OC curve decreases with such deviations, and

also importantly, this is achieved at the expense of eroding consumer protection as can be seen by these curves having a pretty high probability to accept pretty bad batches.

Now, we are not complaining that these deviations have been allowed because they have been necessary and well justified; otherwise they would not have been approved. What we are saying is that this demonstrates that the capability of many products is not such that they can live with the current draft guidance curve.

Now, the PTI test provides a comparable reduction of consumer risk without compromising consumer protection, demonstrated by the fact that producer risk is reduced, whereas consumer protection is maintained.

As I said before, the point is that fewer rejections does not necessarily mean lower quality of accepted batches. I will demonstrate that by showing you two cases of simulated or computer-simulated situations, one for unacceptable quality, where I'll show that the FDA and the PTI test have comparable performance in consumer protection, and the other case is for acceptable quality, where I'll show that the PTI test rejects fewer acceptable batches than the FDA test, yet the quality of those accepted batches are virtually the same.

Now, this is a busy slide. I'll try to explain it to you. First of all, each of the panels show batch standard deviation versus batch mean, and each dot on each panel represents a batch with a true standard deviation and mean as merited by its placement on this panel. The upper two panels are for the FDA test, the lower panels are for the PTI test. Panels on the left are for batches. It's the quality of the batches that were accepted by the test. The panels on the right depicts the quality of the batches that were rejected by the test. As you can see, the batch mean and standard deviation vary here, and they vary approximately for the batch mean between a 100 plus/minus 14 percent label claim, for the batch standard deviation approximately 20 plus/minus 3 percent standard deviation.

The take-home message on this plot is that with this unacceptable quality, the FDA test and the PTI test do a good job of rejecting the absolute majority of these batches, and this just further illustrates my point that the PTI test achieves the goal to maintain consumer protection.

The next panel here, which is also a very important slide, shows the case for acceptable quality. So you can see here from the left panels that with the FDA test, 65 percent of these hypothetical simulated batches were accepted, whereas with the PTI test, 95 percent of the

batches were accepted, yet the coverage of these accepted batches is virtually the same at about 98 percent coverage.

Now, take a look here at the quality of the batches accepted by the FDA test and those rejected by the FDA test, and you will see that the quality is not that much different; whereas with the PTI test, there is a clear distinction in quality between those accepted by the test and those rejected by the test. Now, this is due to the PTI test having a steeper OC curve being more efficient in discriminating between quality.

I'd also like to point out that with the 35 percent of the batches rejected by the FDA test, as you can see, this does not necessarily mean that the high rejection rate figure here, 35 percent, that these batches have been rejected due to poor quality. Most of these batches have been rejected by the test because the test is not very discriminatory. So it's a feature of the test that gives you the high reject rate. These illustrations show that the gap is of lower relevance than perceived initially from the OC curves.

Now, let's move back from producer risk assessments to consumer protection and quality standard. We firmly believe that quality of a batch should be judged against a specific standard. Within the presented hypothesis framework, that standard is the limiting

quality, defined as the quality corresponding to 5 percent acceptance probability; that is, a high confidence of rejecting such a batch at the limiting quality. This addresses consumer protection issues, and as I said, a typical batch quality has to far exceed this quality to achieve reasonable acceptance rates.

A quality standard should not be simply a decision rule based on some typical batch quality. That would not provide the hypothesis regarding what is acceptable or unacceptable quality in a batch. That is simply a decision rule which is completely inflexible and completely tied to the sample size on which this decision rule is based. So there is no flexibility.

It also would not be simple to cater for different products having different typical qualities. There would be no mechanism, except to make exceptions from the decision rule to cater for such a situation.

As you remember, the proposal is that the limiting quality is set to 85 percent coverage of the 75 to 125 percent label claim interval, and this is the same limiting quality as implied by the draft guidances. And as you remember, this should be demonstrated for each batch with high confidence.

FDA has commented that a tighter standard may be needed. We argue that a significantly tighter standard

will be problematic. A standard must be compatible with the capability of products it is regulating. So it has to be commensurate with the capability of current and pipeline products and with the associated analytical methodology, and in setting that standard, both producer risk and consumer protection should be considered. If the standard were to exceed capability, that would create difficulties for manufacturing and especially for development and approval of new products and generic versions.

Now I'm going to talk about normal distributions and zero tolerance criterion, also one of the issues raised by the agency.

The statistics of the PTI test is based on normal distribution. We have a database collected that demonstrates that this assumption of normality is appropriate. To challenge the test, though, we have studied a number of non-normal distributions and recently non-normal distributions that have been suggested by the agency to be very challenging non-normal distributions.

Our investigations have revealed that with the revised PTI test coefficients, the PTI test assures less than 5.1 percent type I error at the limiting quality for all normal and for most non-normal situations. For a few extreme distributions, 5 percent is exceeded at the limiting quality. These extreme distributions are not

reflective of real products. They are significantly off-target, relatively symmetric distributions with extremely short tails or they could also be significantly off-target, notably asymmetric distributions with the longer tail in the off-target direction. Now, we conclude that the PTI test is appropriate for real products.

Zero tolerance has also been a criterion, mostly because it's part of the present guideline test. It has been under consideration whether or not the addition of a zero tolerance criterion through the PTI test would be a benefit or not.

A fixed zero tolerance criterion has been shown to degrade parametric tests, and this effect escalates with the sample size. This is simply due to the fact that if you introduce a nonparametric criterion, such as a zero tolerance criterion to a parametric test, that will convert the test from being parametric to being nonparametric and you will lose the efficiency.

So a zero tolerance criterion must scale with the sample size in order to avoid degrading the parametric test and to have no effect on producer risk. We have shown that such a scaled zero tolerance criterion has little or no effect on consumer protection, even for the most extreme non-normal distributions. So our conclusion is that zero tolerance does not help control product quality.

And this is just to illustrate my point and we can look at the lower row here. First, I'll tell you that this is the acceptance rate at the limiting quality, the 85 percent coverage. So the acceptance rate figures are given here with the zero tolerance criterion and without the zero tolerance criterion, and this is for the small test, same thing with the big PTI test.

Now, we can take the most extreme non-normal case which is the asymmetric short-tailed beta distribution with alpha equal to 2, beta equal to 100, off target at the worst position. We see, as I've told you, that the acceptance rate exceeds the ideal 5 percent, but we can also see that the addition of this problematic zero tolerance criterion doesn't really materially improve this consumer protection. So the conclusion still is that zero tolerance is not helpful in product quality assessment.

Now, I've given you the overview with focus on most of the issues, such as revising the coefficients to make true the 5 percent error rate. I've discussed the quality standard, the perceived gap between the FDA and the PTI OC curve, issues about non-normality and zero tolerance criterion, and I hope that we can all agree that the PTI test is conceptually acceptable as a replacement, parametric without the zero tolerance criterion and with coverage as the quality definition.

A desirable characteristic of the test is that it allows product-by-product justification of the sample size, and this, with the same consumer protection, but this is then the mechanism to mitigate producer risk while maintaining consumer protection at a constant level. And this consumer protection then is that implied by the FDA guidance test.

Thank you for your patience.

DR. KIBBE: Is there anybody who would like to ask a few questions? Efraim?

DR. SHEK: Just a clarity. We were talking about that this product is a combination of the formulation and a device. Those proposed tests, do they de-couple both of them? Because you have an actuator, you have a pump and other devices, and that might be the same for all the rest, whether it's the guidance or what you're proposing.

DR. OLSSON: No, they do not de-couple the performance of the device and the formulation. These are tested as a unit, as is appropriate, because that is what the patient experiences.

DR. SHEK: But we might have different batches. Let's say the actuator is being made and you are using it for various batches of the canister. So we'll repeat those testing, I would assume, and we assume that the actuator passed as a batch.

DR. OLSSON: Yes, that is a complication, and as we've said, the test we are now talking about is only one of a number of strategies and tests used in order to ensure quality products.

DR. KIBBE: Thank you.

We have another presentation already to go. Dr. Hauck?

DR. HAUCK: So good morning. Two largely statistical talks in a row, so hang in there. This makes it a tough morning for you.

I should also say that there's a certain amount of overlap between the three presentations that you're hearing, and given that we're not tight on time, I'm going to go ahead and sort of proceed as if the overlap is not there and hoping that the same things from three different perspectives will be helpful to you rather than just boring. So let's see how it goes.

So I was asked to assess the IPAC-RS proposal, and I should say that the slides had to be made up prior to the receipt of their recent report. So this is largely based on their 2001 proposal and I'll try to remember to indicate, as I go through it, how things have been changed, based on the most recent report and the presentation that you just heard.

So I'm going to look at some of the issues that

have been raised regarding the FDA draft guidance and how the IPAC-RS proposal addresses those issues and then my own view as to whether the details of what the IPAC-RS proposes support the claims that they make.

So this is the FDA proposal. You've seen it a couple of times. It's what's called a two-tier or a two-stage testing proposal: first tier, 10 containers with acceptance criteria that I don't need to repeat. It goes to the second tier, an additional 20 containers for 30 total, and then criterion at the second stage. And as has been mentioned a couple of times, we've got a requirement on the mean and a zero tolerance criterion at both stages.

So really there are three pieces, as Dr. Adams had alluded to. We've got an inner interval, which is sort of the formal test by attributes as the quality control language uses it; the outer interval, the zero tolerance criterion, sometimes referred to as the safety net; and then the limit on the sample mean.

So one of the issues that has been raised regarding the FDA proposal is in front of you. The idea is that what you're looking at is something that very much looks like a statistical hypothesis test. You collect some data. You perform a statistic. If the statistic satisfies certain criteria, you say pass, and if it doesn't, you say fail, and the only thing missing from it is the hypothesis.

So there's no statement of what constitutes an acceptable batch, and this is what Dr. Adams was referring to.

So the focus, in effect, on the FDA proposal has been on what's an acceptable sample and not on what's an acceptable batch. And I think the original issue raised is to say that seems kind of backwards and inappropriate, that the FDA's role should be to specify what's an acceptable batch and then the sponsor should then get to decide what sample they want to take.

So what IPAC-RS does is it essentially accepted that challenge and, as you've been hearing, they set down a specification referred to in the two previous talks as the limiting quality standard. They propose the 85 percent of the batch falling within 75-125 percent of label claim, a number that they obtained by evaluating what the FDA proposal actually was doing.

Now, what that looks like is the following. So this is again, as Dr. Olsson had alluded to, based on normal distribution, and it's intended to show you the combinations of means and standard deviations in the batch that correspond to that limiting quality specification. Remember, that's 85 percent of the batch falling within 75 and 125 percent.

So the idea is that anything that's inside that red line should be acceptable because anything inside the

red line satisfies the standard of at least 85 percent of the batch falling within 75 to 125 percent of label claim.

Then that gets us into some of the language that you've been hearing already this morning and I'll elaborate a little bit. The term sometimes is called consumer risk. We can call it a false positive. Also the statistical language would be the type I, or alpha error, and that here in this context means that a batch that lies outside the specifications; that is, any batch that lies outside or above the red line here, the probability that that batch would actually pass whatever the rule ends up being. And the producer risk, the converse of that, is that a batch that falls under the red line and meets the specification set, the probability that that batch fails. Now, normally in trying to design studies, we always like both those probabilities to be small. That's always the goal of study design.

Now, I should mention here that, as has been alluded to, that the issue of what that limiting quality should be is clearly on the table, and the type I/type II errors are going to very strongly depend on what that choice is. Just to give you a bit of a flavor for that, Dr. Adams had put up some of the alternate choices that at least conceptually could be considered, and this just shows you what happens to that definition of acceptable batches,

if you tighten up the IPAC-RS proposal which is the red line going down to keeping the 75-125 but changing the content to 90 percent, which is your green curve, or keeping the content but tightening the limit, the blue curve. You can see a pretty substantial drop, particularly here at the top, in terms of the variability.

So a couple of different comments. First of all, when you're approaching it as the IPAC-RS has in setting an acceptance criteria, whether it's this particular acceptance criterion or some other number, what you're really saying is anything that falls inside that curve should be acceptable. Now, this is really like bioequivalence, if you want to go back to that. You set a limit of 80-125. That really means that 124 would be acceptable.

Now, in practice, you're not going to see that because the size of the study required to get something 124 to pass would be unreasonable. So you're never going to actually see that and actually Dr. Olsson really alluded to that in a different way by indicating that batches really need to be substantially inside that red curve in order to have a good chance of passing.

The other comment to make is really relating to the second question that Wally had put up for this committee, and that's really that if you take this approach

and say that the role of FDA is essentially to set the stake in the ground and set the quality limit, then the batch failure rate for batches that are acceptable really becomes the problem of the sponsor and not the problem of the agency because they get to choose the sample size.

I thought it might be useful for you to see a little bit of the difference between what -- I think there sometimes seems to be confusion going back and forth between batch and sample. So this is just intended to highlight for you, the red curve again being what defines an acceptable batch and the green -- I guess I'll call it a curve but a pentagon or hexagon there -- is samples that satisfy the 2001 IPAC proposal of sample size of 30. You can see it's substantially inside the red curve.

The second issue I wanted to talk about that has been raised regarding the draft guidance is that the FDA is fixing the sample size and any time the regulatory agency fixes the sample size, it's really denying the sponsor an opportunity to control their own producer risk.

So the IPAC-RS proposal does provide a choice of two-tier designs and, as you heard from Dr. Olsson's presentation, all of which are intended to maintain the false positive rate of 5 percent for each possible sample size they consider. This is sort of a personal opinion. There's nothing special about two-tier designs. There's

certainly nothing special about the two tiers having one-third on the first tier and two-thirds on the second tier. That seems to have historical significance but no real scientific significance, and there's really no reason why the number of tiers and how they're split up can't be variable as well.

And again, this goes back really to a prior comment. As long as the batch meets the specification of what's an acceptable batch, then any sample size should be acceptable, and this really kind of is going to raise some issues, I guess, that you'll have heard already, and that part I've covered.

The third issue that is raised regarding the FDA proposal, and really a bunch of other proposals, is that by using a test by attributes, again the quality control language, it's making inefficient use of the data. Now, that's sort of statistical language meaning that your precision or your hypothesis testing is not being done as well as it could because you're not making best use of the data, and so it moves to parametric tolerance intervals. As Dr. Adams indicated, this really originally was based on the JP proposal. It does eliminate or rather reduce some statistical conservatism that's present with tolerance intervals, and I'll show you a picture of what that looks like in a little bit.

The fourth issue I wanted to talk about is the zero tolerance criterion. You've heard that quite a bit already. I think one of the main points there is that there's really a complete disconnect or conflict, if you will, between having a zero tolerance criterion and allowing variable sample sizes because you're really saying that this is something that you're going to have to fail the larger and larger the sample size gets, even without getting into the multiplicity of times that the test is done over the course of a year for the companies.

So now that I've seen the new work particularly, I think I could even go a little further in a personal opinion and say that somebody who wants to argue for zero tolerance criterion really has the burden of proof on them at this point.

So as you did here, the IPAC proposal does drop it, and the last point was really made for me prior to this morning. I think the zero tolerance criterion certainly does seem to engender a level of comfort and I don't know whether or not that comfort will still be there or whether there's enough data at this point to make people comfortable about dropping it. Clearly, as I'm saying, in my opinion, that's really the way to be going.

So summarizing the different issues, so yes, I'm saying I agree that the IPAC proposal does address the

issues that have been raised about the FDA draft guidance. And as I add here, and I think Dr. Adams alluded to, although we're talking about the FDA draft guidance, there's actually a number of other proposals that have been out there prior to this and what we're talking about for the FDA applies to them as well.

Now, one of the major claims in the IPAC-RS proposal is at the same time as you can maintain the level of consumer risk or even improve the level of consumer risk compared to the FDA criterion, you can still reduce producer risk. So this is one of those "how's that possible" sorts of thing because normally in study design, you think of those two things as being trade-offs. You can have difficulties doing both at the same time.

So first of all, this is actually the difference. Again, this is now the 2001 criterion, first stage with an n of 24, so this is out of their report. So the blue curve shows you what would be acceptable samples, based on a standard tolerance interval approach, and then the green curve is the IPAC-RS proposal. You saw me put a limit on the sample standard deviation. That's what puts the flat top. In exchange for that, to maintain the 5 percent, they get to add some shoulders to the curve, and then the red lines, which hopefully show up, are the plus or minus 15 percent on the sample mean.

So how are they able to deliver? So part of it's the parametric to nonparametric difference. The parametric approach will give you lower producer risk for a given level of sample size and consumer risk. So that's part of it. The elimination of zero tolerance criterion is certainly part of it, and then I bolded the last one because one of the things that -- I don't know if you noticed going by there, the sample sizes in the IPAC-RS proposal are largely larger than were in the FDA proposal, and there's no question that that's going to be part of the package in terms of making it possible to do what they're planning.

I also should mention that the FDA's draft proposal is more liberal than it appears. Remember, in Dr. Olsson's presentation, it came up that the implicit limiting quality standard in the FDA proposal is either 75 percent or 85 percent coverage within the 75-125. That wasn't in the FDA proposal and this is sort of a reverse engineering issue because, remember, there was no proposal on that. So I think that was more liberal than it was expected.

So I think I'd summarize this part of it by saying that yes, the IPAC-RS report does deliver as claimed on this, that this is an improvement in statistical methodology. The only thing added here is you need to be

careful in the choice of the constants. I think you've heard some of that already from both Dr. Adams and Dr. Olsson.

This sort of weird picture is to give you an idea of what's going on. If we had an ideal test here, we'd have the magenta line in the center. So this is looking at the first tier actually of the two-tier test with 24 samples in the first tier, and the target here is 2.5 percent consumer risk, not 5. So ideally, we'd have 2.5 percent all the way along as the mean goes from 75 percent to 125 percent of label claim. The blue curve shows you the old standard normal theory of tolerance intervals, what it does. It's nicely right in the center but drops off very quickly. And I had alluded earlier to statistical conservatism in the normal theory tolerance intervals and that's what that is. This gap between the blue and the purple is something you'd like to do away with and because you're increasing the producer risk here by making the consumer risk less than it needs to be.

Now, the problem is that, depending on your choice of constants, using the approach of the IPAC, you can end up with things that look like this, so this goes, instead of the 2.5 percent where it should be, up above 4 percent here and coming back down. Now, again remember, this was all based on the 2001 report and they've changed

their constants since then.

I did want to throw this one in here just to assure you that the issue of maintaining the level of risk is not the structure or the form of the IPAC-RS proposal. It really is just an issue of the choice of constants. So there's really, if you will, a dose-dependency here. You can change the constants. They have what they call the f factor which limits the standard deviation. So we have .8 here. This is the IPAC proposal which has a value of just under .8 at the time, .8, coming down to .9, and then coming down to 1 which is the original regular tolerance interval. I think you can see it's really just an issue of picking the constants right to maintain things appropriately.

So I thought I'd also summarize in terms of the IPAC proposal in terms of cost to sponsors because it's not all plus-plus. It's not, you know, just gravy there, if you will. First of all, as I indicated, for the most part, the sample sizes are going to be large, so there is an increased cost in that respect, and the details for the multi-dose products, there is a reduction in cost because rather than testing beginning, middle and end separately, it's combined into a single criterion. And then the biggest cost is not passing when you should pass, and so potentially by giving sponsors control of their study

design and hence of their producer risk, there's at least potential reduction in cost there as well.

So the bottom line really goes back to where I started in the first issue. The message, if you will, is to not spend time on the statistical issues. At the end of the day, take the statisticians, throw us in a room. If we ever agree on anything, let us out. That might be a long meeting. But I think the primary issue for this committee and for the FDA is really what's the limiting target. What is an acceptable batch once you get to market? And as I said, that's my bottom line for you.

Thank you.

DR. KIBBE: Questions?

DR. KORCZYNSKI: Just so I understand the topic a little better, we've heard of dose uniformity here in these presentations. What's the relationship to aerosol particle size? Because that would influence the availability of the drug relative to uptake by the respiratory system. Is that an independent variable? Is that measured in a separate set of tests, and is that considered in any way related to dose uniformity?

DR. HAUCK: Well, I think I should turn that one over to Wally.

DR. ADAMS: Yes. That's a good question. We're talking here about dose content uniformity testing,

and we are not talking about delivery of the drug to the pulmonary tract or to the nasal passages.

This test is based upon the drug ex actuator; that is, after it's fired, it's the dose of active drug that is emitted from the actuator, independent of particle size distribution and independent of delivery to the lungs.

Does that answer the question?

DR. KORCZYNSKI: Yes.

DR. KIBBE: When you eliminate the zero tolerance and you do a statistical analysis, at what point would the batch fail? For instance, if one sample out of a group of samples that were taken, one item had absolutely no material in it, statistically, that might still allow the numbers to come out such that the batch could pass, but I would wonder whether there would be some remedies taken within the company to find out why it was completely empty. At what point do you start to make, I don't know, decisions that go past just the strict adherence to the test?

DR. HUSSAIN: Let me try to put another layer of issues here, and I think, as I listened to the presentations, I think it came across as if this is a final test. I would like to sort of remind the committee that I think as we develop your product, as you go through your validation, all these essentially are addressed. In routine production, it's not a hypothesis test. The

hypothesis test essentially has occurred in terms of development and validation, and I think the confirmation that you have during routine production is simply making sure you're reproducing your validated products.

Now, going back to sort of the issue, Art, you raised, I think today, for example, when we use a zero tolerance criterion, when we reject a batch or when we accept a batch, often, sometimes, there's no difference in the batch quality. It was simply a statistical -- even that sort of triggers that, and I think that's the point that was being made.

I think what this proposal does is to enhance the science of manufacturing from a validation perspective. I think, from development to validation runs, you bring variability as an additional measure of your process capability. It sort of opens that door for that analysis, and if you really look at it, as you go through the validation runs, when you start determining whether your samples collected are normally distributed or not, that I think tremendously helps to make sure the samples we collect later on during validation are more representative and actually could be focused on where the high risk might be. And you can take this back and connect it to, for example, the PQRI blend uniformity proposal that went for stratified sampling. So I think that's the part I wanted to

make sure we understand.

DR. KIBBE: This individual test has to then have additional requirements on when the samples are collected during the run and what happens if there are blanks.

DR. HUSSAIN: That's the point I want to emphasize, is process validation is planned to address that. I think we have not discussed that or presented that part of the work to this committee. It was simply focused on the statistical criteria, but there are layers and layers of approaches and then work that is done to eliminate that possibility.

MR. SCHUIRMANN: I just wanted to add that looking now just at the dose content uniformity test as opposed to the whole battery of procedures that need to happen for a batch to be released, for the small version of the proposed test, 10 samples in the first tier and then 20 additional if you go to the second tier, if there were a single dose with zero content, then it would be impossible to pass the test, regardless of how the other observations came.

Now, I think this calculation could be done, and I apologize, I haven't done it. If the sample sizes were larger, there could be a large enough set of sample sizes that there could be a single zero and it would still

pass. I can't tell you what that is. I think the sample sizes would be very large indeed.

DR. KIBBE: Jurgen?

DR. VENITZ: I have two questions. One is probably a stupid one, but what does IPAC-RS stand for?

DR. ADAMS: It stands for International Pharmaceutical Aerosols Consortium for Regulation and Science.

DR. VENITZ: Okay. Thanks.

The second one may be a more intelligent question. It relates to the PTI mechanics and that's a question for Dr. Hauck or Dr. Olsson.

I'm working my way through the algorithm, I guess, and it sounds like one of the predefined things that has to happen prior to doing any of this is to agree what those k and f values are. In other words, that's not something that the sponsor prespecified, but that's something that would be part of a guidance because that defines how your alpha distribution looks like relative to the ideal test. Is that correct? Because it sounds like IPAC-RS changed those constants to make the test more amenable.

DR. OLSSON: Those test coefficients are the essential motor of the test, so to speak. So what one does is to carefully calculate before what those test

coefficients should be in order to give the test the desired characteristics.

So yes, those are predefined and it's a lot of work to calculate them. We've calculated them for a number of sample sizes to give this desired coverage of 85 percent as the limiting quality. If that were to change, then it would be different coefficients.

DR. VENITZ: So I think what you're saying then, if you assume that you want to maintain the 85 percent coverage, 75 to 125, then the only other piece of information that you need is a sample size and then you can calculate the k and the f?

DR. OLSSON: Well, we already have done that. So they are already in the public arena.

DR. VENITZ: Right. But they would be then part of some guidance if this ultimately evolves into a guidance?

DR. OLSSON: I would believe so, yes.

DR. KIBBE: Gary?

DR. HOLLENBECK: First, I'd like to thank everyone for their presentation. That was very informative. I think we talk about science-based regulatory policy. If you ever wanted to point to an example, I think this is a very powerful one.

I'll also ask a couple of stupid questions, I

think, here.

First of all, it seems that the 5 percent alpha is a fixed given, and I didn't hear a lot of discussion about that. How is that number arrived at? What goes into the thinking that says that's an appropriate level for consumer protection?

MR. SCHUIRMANN: Well, I think that it's mainly a matter of tradition. There are a number of FDA testing procedures that have adopted 5 percent as what's called level of significance, maximum tolerable chance of approving something that shouldn't be approved. There are some other situations in FDA regulations where the de facto level of consumer protection is 2.5 percent. There certainly could be and probably have been arguments that that's what we should be using here.

Certainly if discussions led to the assertion that not 5 percent but 2.5 percent, or any other number you would care to specify, is the appropriate level of consumer protection, then IPAC-RS could have reverse engineered the FDA proposal, found out what level of quality has a 2.5 percent chance of being approved and called that the limiting quality and designed their test to assure that same limiting quality and all those things could be done.

But 5 percent has been a traditional level of consumer protection. It's thought approving something

unacceptable 5 percent of the time, I suppose, is rare enough that it's not a concern but not so very rare that in order to assure it, you have to do arduous testing, but certainly that number is at the discretion of the regulators. It doesn't come from the statisticians.

DR. SWADENER: The 5 percent also has not only been tradition in FDA or those kind of circles. It's other fields as well. I'm from education and it's very common in that field.

DR. DeLUCA: Yes. I noticed when Wally had some questions, Dr. Adams, you also listed as one of the options 90 percent of the doses within 80 to 120 percent. Walter, in your treatment, that option wasn't included. The other three were. I guess that's one question. What would the treatment look like if you included that?

And then, I guess the rationale for not maintaining the sample mean of 85 to 115 percent -- that was part of, I guess, the FDA draft and the PTI, and I'm wondering why that was not maintained. So I don't know what the treatment would look like if you included these two options in there.

DR. KIBBE: Is there an answer?

DR. HAUCK: Yes. I was trying to find my copy of my handouts so I could take you through. If I remember right, the 90 percent within 80-120 was, I think, more

strict than anything I've put up there. So the curve would be -- no. It's earlier.

Now, I wasn't -- I didn't have the 85 thing on the mean in most of what I showed because I was talking about the criterion on the batch and the plus or minus 15 percent is just a criterion in the sample. So if you look at the slide 12, the vertical bars on the right side of the green -- well, it was green on the hexagon in the lower piece. Those vertical bars are the plus or minus 15 percent on the sample mean. So as long as that piece is in the criterion on the sample, no matter what the sample size or anything else, you'll have those vertical bars at 85 and 115, but that's on the sample. It's not a batch criterion.

Now, back to your first part of your question, on slide 10, 90 percent within 80-120 would be under the blue curve. It would match in the corners. The blue curve is the bottom of the three curves. So it would match in the corners but be lower in the center.

DR. ADAMS: Dr. DeLuca, in addition to that, the 90 percent within 81 to 120 was a more recent suggestion that we had come up with subsequent to Dr. Hauck preparing his slides. It gets to this issue of what Dr. Olsson called the gap which is the distance, the difference in standard deviation at a particular probability level between the FDA curve and the IPAC-RS curve and an interest

on our part in trying to possibly move that operating characteristic curve for the IPAC-RS to the left, reducing the gap, and that's where that number was suggested.

DR. SADEE: I can see the value of not having the limitation for production purposes. On the other hand, I think that my question would be what is the risk of incurring an adverse event? 5 percent, for instance, would be unacceptable. So if one goes further and further out of the range, then at what point is there a risk of an adverse reaction? If there's too little in a metered inhalation, then there might be a second dose taken by the patient. That might lead to an overdose. If it's too much, it might lead to an overdose. So I think what one should factor in is a statistical analysis of the risk of adverse effects and that should determine where there is a limit.

DR. ADAMS: May I comment on that? Just one thought is that with regard to the variability in the products, and you've mentioned about multiple dosing, patient taking multiple doses, you know, one consideration might be, while we are talking here in the context of a single standard across different dosage forms, MDIs, DPIs, nasal sprays, and across all drug classes, that different standards conceivably could be appropriate for, let's say, an inhaled corticosteroid than for a beta agonist where, with the beta agonist being used as rescue medication is

important that that drug product on a given dose to deliver the expected dose. Possibly on a chronically administered product, maybe greater variability could be allowed, but at this point, we have not made such considerations.

DR. HUSSAIN: I think the question is the right one, but I think the answer, I think I would like to sort of propose is, what happens today and what happens with the current FDA test and what happens with the PTIT? There's no difference.

If there is a canister which is 0, has not content in it, what is the probability of finding that with the small sample size that we test today? When it happens with the PTI test, it's going to be caught anyway. I just want to have Don explain that a bit more.

MR. SCHUIRMANN: Well, there's nothing much more to explain. Dr. Hussain is particularly talking about a zero content canister, one that somehow didn't get any drug in it. I assume that the adverse reaction you're worried about would come on the opposite end of the spectrum of it has too much in it.

If there's a canister lurking out there with 200 percent of label claim in it, the chance that it will end up in your tested sample is the same, no matter whose test you're using, the FDA draft test or the proposed parametric test. If a canister with 200 percent of label

claim actually did show up in the sample, I suspect that it would cause either test to reject the batch.

Now, I've picked 200 percent out of the air. We could play with the numbers and you could eventually come to an amount where the zero tolerance feature would kick out that batch, but the parametric test would let it pass, and then the question is, if the content is low enough that you're in that zone, is that the type of content that would lead to an adverse reaction, and that's not something I can answer.

DR. HUSSAIN: I think the point I'm making here is, I think, the thought process that this is a test. This isn't a production run. How representative is the sample, first of all, because you're testing a number of small samples to just make a decision. What I would argue is, I think, a parametric approach, a more rigorous statistical approach reduces the risk of that happening from the current situation and the reason for that is, I think you are using the information more scientifically because you understand your variability, you understand the distribution of your material which we may not be doing today.

DR. SADEE: Yes, but we do have to consider the risk for each individual drug which is very different. If there's a therapeutic index that's very narrow, then you

have to --

DR. HUSSAIN: Definitely.

DR. SADEE: -- be much more stringent, so we cannot talk about one standard. You have to reflect that.

DR. HUSSAIN: No. That's a very good question, but I think when we talk about two different approaches, I think you have to look at how is this approach protecting that and how is that other approach protecting that. What I would sort of suggest is with a rigorous statistical basis, the proposed test would protect it better. So that's the point.

DR. KIBBE: You had something to say?

DR. TSONG: Yes. I just had prepared two slides to address the general issue of quality standard. Could I show them?

DR. KIBBE: Fine.

DR. TSONG: First, I want to get permission from Dr. Olsson because I used your slide and twisted it a little bit to get to my point.

(Laughter.)

DR. TSONG: First, let's talk about the quality standards. Suppose I'm a drug manufacturer and I have a supplier to supply the material, and so whenever the shipment comes, I have to take a sample to give a quality score of that. Suppose the perfect score is 100, and once

I receive a product which scores 75, I know I'm going to reject the batch, turn it back.

But the chances are I receive a batch which has a score of 90, which I feel, hey, I could cheat a little bit, so I'll pick up the phone and call the supplier, and say this quality is not really what I expected. I wanted 100, you gave me 90 percent, and you have to pull your act together to give me better product.

Then if I get a score which is 85, and I probably would tell him from now on every 10th batch, I'm going to reject one of them, turn it back to you as sort of a penalty. I don't need to get a complaint from my customers. So this is a 10 percent rejection which also plays a role in the quality control there.

So we have a couple of points there. One is the minimum quality, one is the quality assurance I wanted to have the product to be. So in setting about a quality control procedure, we need to take both of them into consideration.

Now, I wanted to show you this slide here. This is slide 4 from Dr. Olsson. Here, it shows that at the lower right-hand which controls the type I, 5 percent type I error rate, which is the consumer protection region but really what it means is it's a not acceptable batch which really we don't want this kind of batch to be

released. And on the right-hand side, it has the producer protection region, but this region is going to be changed with the sample size. If the sample size increases, this region can be shifted up to here. That means many of the batches of the area of uncertainty, which it really means for the consumer which is the product, is not totally bad. It's not as totally good as we want. So that means if the sample size increases, many of the uncertain quality batches can be released.

So what do we really want to consider? We have to consider the quality assurance region, which means I want the batch to be of this quality, and if it's below this, I'm starting to reject the batch with, say here, 10 percent of rejection. If it's worse than that, I'm going to reject more.

So we need to fix the level to have a good quality control. That's what is question 2 of Dr. Wallace presentation, what is 10 percent, and I think that's the 10 percent interpretation for quality assurance.

Now, we have the discussion and some of those iterations are how we going to set up this point. I think the original one is this one. We have this as, say, that's original FDA procedure. You have 10 percent rejection at this point which is about 9 percent of the standard deviation. And that's what is suggested. Probably we need

to start looking at this point for the quality assurance region.

And the gap here does bother us. The longer the gap, that means the further away from assurance quality, and with the sample size increasing, you have higher protection for the producer risk, but you have less protection for the consumer of those marginal quality products, and I think that's a point regarding to the two questions.

What I'm trying to say is that we are not questioning the quality limiting approach, but we are setting up the question, what is the standard we want to put out for the setting up the quality control procedure?

DR. KIBBE: Does anybody else have a comment?

MR. SCHUIRMANN: Just to expand on what Dr. Tsong was saying. Suppose that I'm a product manufacturer and I have a process that tends to produce batches of metered-dose inhalers that over the whole batch average about a 100 percent of label claim. My process is on target, and my process tends to produce batches that have about a standard deviation of 11. 11 what? 11 percentage points of label claim. So that's the measure of variability of the delivered dose from individual actuations of my product.

Well, if I start producing lots of batches and

applying the FDA test as described in the guidance, I'm going to only be approving about a little more than 65 percent of my batches. 35 percent of my batches are going to be rejected, and as Dr. Adams mentioned, the court decision would lead that to be taken as evidence that my process isn't in proper control.

On the other hand, if I apply the proposed parametric test, I'm going to be accepting more than 95 percent of my batches, based on this test. Now, as has been often mentioned, there are more than one test that gets done to a batch before it goes out the door, and this test isn't necessarily the gatekeeper.

But still, in my hypothetical example of batches that tend to have a standard deviation of 11, I'm going to accept most of my batches and release them, based on this test, using the proposed test, but I'm going to be rejecting an unacceptable percentage of my batches if I use the FDA test.

The issue is that the FDA test is doing the wrong thing and the proposed test is doing the right thing, if a batch of standard deviation 11 is acceptable to the public health. On the other hand, the FDA test is doing the right thing and the proposed test is doing the wrong thing, if a standard deviation of 11 is not acceptable to the public health.

So we've already heard talk about the limiting quality; that is, defining the batch that anyone would agree is an unacceptable batch, but we somehow need to define an additional value which is the quality, the level of quality that corresponds to, if that's routinely accepted, that's a good thing.

I might point out, also, say I have a process that produces a standard deviation of 13. Well, now, with that process, the FDA test is going to be accepting fewer than 50 percent of my batches. Similarly, the parametric test is only going to be accepting about 62-63 percent of my batches. So in either case, I'm in trouble, but this curve, this blue curve is for the proposed test with 12 in the first tier and an additional 24 if you go to the second tier, but if I increase my sample size, I can make the operating characteristic curve for the proposed test go higher and by taking a large enough sample size, I can make it go higher than 90.

So the issue that is currently occupying our attention in CDER is whether we need to specify this additional level of quality to be assured and how can that be done.

DR. KIBBE: Thank you.

DR. HAUCK: If acceptable, I wanted to go back briefly to the question raised about the empty canister and

the zero tolerance criterion.

DR. KIBBE: Sure. Enjoy yourself.

DR. HAUCK: The problem with the zero tolerance criterion in the FDA draft proposal is it really impinges on normal variability. That's what makes it sort of a guaranteed to fail sort of thing eventually. You can imagine setting -- I should put a different name on it. You can imagine setting some sort of, say, clinically acceptable limits or some much wider than that, saying if there really was a canister that had 10 percent in it or 300 percent in it, that we don't want that to be in a consumer's hands, and if by some stroke of luck that should show up in a sample, that would be a problem. It would be a much wider type of zero tolerance and that sort of thing would probably not impinge on the producer risk in terms of normal variability.

DR. KIBBE: Anybody else? Gary?

DR. HOLLENBECK: Is there a concern when the distribution is not normal? Whoever would like to respond.

DR. HAUCK: Yes and no, I guess. You've got four statisticians in the room, so you'll get 15 different opinions on this one.

Normal theory tolerance intervals can be a problem if you deviate too far from normality and that's what you just saw in Dr. Olsson's presentation, and so we

always know when we do parametric methods that you can find some situation that makes it a bad thing to do, but you then have to ask, well, what situations are reasonable and plausible to worry about here, and that part of it, I can't answer. I could turn it over to Don and Bo at that point. And then you'd want reasonable confidence that the alpha level is at least close to 5 percent on reasonable, plausible alternatives to normality.

DR. TSONG: Could I answer the question, too? I think if we go back to the original one, which is the statistical paper that proposed the tolerance limit, I think currently used and maybe a little bit modified -- but the original work shows that if you use the original tolerance limit, that really the approach is slightly conservative, which means if we say 5 percent have whatever rate, when you calculate out, it's really lower than 5 percent. That means you release less than 5 percent for those you're supposed to release 5 percent. There's also lots of work done that shows that if it's not under normal distribution, what is going to happen.

I think that if it's skewed, if it's skewed, but it's a uni-model, that means only one peak, have a distribution, even when it's skewed, it's pretty much robust on that. But when the distribution is really widely different from normal, that could be totally different.

DR. KIBBE: Ajaz?

DR. HUSSAIN: Just to sort of put an overlayer of an engineering thought process there in a sense, because I do want to link that back to process understanding. If you have a non-normal distribution in your samples and in your content uniformity, now, if that is related to your manufacturing run, is it happening in the beginning of the batch or in the end of the batch, and what is that? I think that provides a level of understanding of process. Is segregation occurring or whatever that mechanism is. And I think this is what allows us to get to the root cause of things and address that because I think the discussion today has been mainly on the statistical aspect of that. I don't think that's a complete picture for discussion.

I think the manufacturing process, understanding the physics of that aspect, has to be sort of brought in. So I think that's the reason we wanted to bring this up as an awareness topic and get your feedback so that we can prepare well when we bring this back again.

DR. KIBBE: Thank you, Ajaz.

I have just a couple of thoughts and that is, the sample size is proposed at 12 and 36, one tier, two tier. That would apply to a batch run of 1,000 samples, a batch run of 10,000, a batch run of a 100,000, and have you looked at the statistical ability to actually detect, with

the same confidence, potential outliers and errors in larger batches with a fixed sampling size?

MR. SCHUIRMANN: It strikes many as counterintuitive, but the performance of the test really doesn't depend much on the size of the batch, unless the the number in your sample starts to become a non-trivial proportion of the number in your manufactured batch. Certainly if you have a batch that has a thousand containers, I would expect it to perform with these tests almost the same as the type of batch that has a half a million containers.

If you had a batch that had a hundred containers, then we might start running up against changes in the performance of the test, owing to the fact that you're sampling a substantial proportion of the batch.

DR. HAUCK: I think the only thing to add to that is that if the batch is sitting out there with any of those sizes, it's got 1 percent or less of some funny unusual values in it, neither of these tests are going to do anything for you and nobody wants to propose a 100 percent destructive sampling which is the only way you'll find it.

DR. KIBBE: We have to start our next little gathering at exactly 11:30 because it is the open public hearing, and we have announced that we would do it at 11:30

and so therefore we will do it at 11:30 Mean Greenwich Time. We're going to check with the Naval Observatory downtown to make sure we're right on 11:30.

So we get a second morning break. Congratulations, everyone. Ajaz is going to take that away from us with a comment.

DR. HUSSAIN: No. Just to wrap up as sort of a conclusion. Conceptually, I think I would guess we would move forward with an in-depth discussion on this and so forth. So you agree with that? Okay.

(Recess.)

DR. KIBBE: I assume that every one of the speakers has checked in with one of the staff and they are ready to go. We hope that we can move through these with a reasonable amount of alacritude, still allowing time for the speaker to say the important stuff that he or she came to say and allowing some of the members of the committee to comment or ask questions, but remembering that we have an hour to get this all done.

I would ask that each speaker identify themselves and the organizations that they are representing or the individuals who have compensated them for their appearance today.

Dr. Wood?

DR. WOOD: I'm Dr. Lawrence Wood. I'm the CEO

and Medical Director of the Thyroid Foundation of America, and I want to acknowledge financial support and in-kind support to help us disseminate our thyroid educational materials and information about the foundation to the patients, the public, and physicians and support for our educational thyroid forums for patients. This support has come from Abbott Laboratories, Jones Pharmaceuticals, Forest Laboratories, EMerck in Europe, and Watson Pharmaceuticals.

The Thyroid Foundation of America is the oldest and largest organization devoted to providing education and support for thyroid patients and increasing public awareness about thyroid issues. We educate our members as well as thousands of others who visit our foundation website that the serum TSH is the most effective and precise way to monitor thyroid hormone therapy. Because of the log linear relationship between thyroid hormone level and TSH, for every 2-fold change in the free thyroxine, the TSH level will change one 100-fold.

Without the reliability and accuracy of TSH measurements, patients with unrecognized hypothyroidism risk complications, including elevation of total and LDL cholesterol, fatigue, depression, decreased work performance, and an overall decrease in their quality of life. Patients with unrecognized hyperthyroidism are at

risk for myocardial infarction, serious cardiac arrhythmias, including atrial fibrillation, anxiety, muscle weakness, diminished productivity, and decreased quality of life.

We're particularly concerned about the importance of TSH measurements in evaluating the effectiveness of thyroxine therapy in patients with thyroid cancer. We must be sure that TSH is fully suppressed to minimize the likelihood of growth and spread of residual tumor throughout the life of these patients. A decrease in thyroxine as small as 12 micrograms can cause dangerous TSH elevations in a formerly suppressed patient. TSH monitoring is also critical since changes in TSH levels can occur due to medications, like iron, amiodarone, Zoloft, and lithium. Patients and even some physicians may not be aware of the potential thyroid effects of some of these drugs.

The FDA has recommended evaluation of thyroid hormone bioequivalence by giving 600 micrograms of thyroxine to healthy volunteers and studying its metabolism by serial measurements of thyroid hormones in the blood. This is inappropriate because it ignores the critical role of TSH in evaluating the bioequivalence of the far more critical tissue effects of thyroid hormones.

We urge the FDA to separately consider this

question with experts in the field of biochemical measurements in thyroid disease.

Thank you for your attention.

DR. KIBBE: Thank you, Dr. Wood.

Our next speaker is Dr. Jacob Robbins.

DR. ROBBINS: I'm Dr. Jacob Robbins. I'm presenting the statement of the American Thyroid Association. I'm Scientist Emeritus at NIH and former President of the association.

The American Thyroid Association is a professional society of 900 U.S. and international physicians and scientists who specialize in research and treatment of thyroid diseases. In fair disclosure, the ATA acknowledges having received unrestricted financial support from companies which produce levothyroxine products, Abbott Labs and Jones-Pharma.

Today's review of bioequivalence for levothyroxine products by the FDA greatly interests the members of the ATA. When L-T4 is used to treat thyroid disease, the patient must receive an accurate and predictable amount of hormone and obtain a reproducible biological effect with each dose. In the clinical setting, the dose is determined by a combination of the presence or absence of thyroid-related symptoms as well as results from thyroid blood tests, especially TSH. Multiple factors

affect the final dose, including body mass, drug absorption and metabolism, the amount of residual functioning thyroid tissue, interference with absorption or metabolism by other medications or food, and patient compliance.

Hormones controlled by a biofeedback mechanism provide a unique situation in which the body provides an indication of whether or not the dosage is appropriate. Close monitoring of TSH concentrations enables practitioners to provide patients with an appropriate amount of medication to ensure that thyroid hormone levels fall within a narrow optimal physiological window.

We understand that bioequivalence for levothyroxine products is currently based on the design which requires the administration of 600 micrograms orally to normal subjects, followed by measurement of thyroxine in the blood over 24 to 96 hours, from which the AUC and the Cmax are determined. For many drugs, this may be very appropriate for determining pharmacologic bioequivalence, acting as a surrogate for therapeutic bioequivalence.

However, in the case of a hormone like thyroxine, pharmacologic bioequivalence only provides part of the story, since absorption is only one component. The biological effect of the medication must also be assessed. Serum TSH provides measurable and critical feedback for assessing the biologic effect of a particular dose of L-T4.

Another important distinguishing factor of L-T4 is the prolonged half-life of approximately one week. Presently, measures of bioequivalence are done after an acute dose, thereby overlooking the time required for hormone equilibration in body tissues. Additionally, one can question the comparability of bioequivalence from a superphysiological dose of L-T4 in a normal person with an intact thyroid versus a patient with reduced or even no endogenous thyroid hormone production. The present technique does not allow discrimination between smaller, more appropriate doses of L-T4.

In summary, in the case of hormone therapy, particularly with oral T4, we have an instance where one can actually measure biological equivalence; that is, the effect on a tissue of the body, which is what bioequivalence should truly mean. Measurement of serum TSH should be done following an appropriate length of time, four to six weeks, to account for the long half-life of L-T4. This would allow the medication's true biological equivalence to be assessed under clinically relevant conditions.

The ATA recognizes the complex nature of the issues being discussed today. Our main interest is to ensure that all L-T4 preparations are reliable sources of thyroxine replacement and that any determination of

bioequivalence for such preparations be based both on pharmacologic and therapeutic bioequivalence. Therefore, we feel it imperative that the biological effect of L-T4 as measured by TSH be part of any method the FDA considers for evaluating equivalency of such preparations.

Thank you.

DR. KIBBE: Thank you, Dr. Robbins.

Our next speaker on the schedule is James Hennessey.

Dr. Hennessey.

DR. HENNESSEY: Thank you. I'm Associate Professor of Medicine at Brown Medical School in Providence, Rhode Island. I've been involved in clinical research with the applications of levothyroxine since 1983, and I have a keen interest in the process to assure that we have reliable and accurate dosing of thyroxine.

I've spoken on this subject at the request of both Forest Pharmaceuticals as well as the Knoll Pharmaceutical, now known as the Abbott Pharmaceutical Company, in the past, but I'm here on my own today, and I've been involved in clinical research protocols sponsored by Knoll, now known as Abbott, and King Pharmaceuticals, in the near future.

At this point in time, L-thyroxine is clinically essential in the treatment of hypothyroidism and

thyrotropin suppression in patients with thyroid cancer, as about 95 percent of those with hypothyroidism have primary hypothyroidism, making the serum TSH a useful and convenient parameter to assure appropriate dose titration. TSH indicates the thyroid hormone action at the tissue level and thus is followed with great attention in the clinical day-to-day management of patients with primary hypothyroidism.

Currently, expert recommended target ranges for TSH in those receiving thyroxine is a very narrow range, between .5 and 2 milli-international units per liter. This reflects the approximation of the currently hypothesized normal TSH range that's seen in the majority of normal individuals. Thyrotropin suppressive therapy with thyroxine in thyroid cancer patients is also considered clinically very useful. Again, TSH is the recommended parameter to follow these patients, but here, the therapeutic window is much narrower.

Recent information indicates that the normal range observed over one year of monthly sampling is much narrower than the range suggested by observations of cross-sectional populations and therefore published in laboratories. In addition to this, each individual demonstrates a unique set point which is their own personal, far-narrower normal range as indicated by the

skew between the patients here.

These observations led the investigators in this particular publication to postulate that TSH values, even within that broadly stated normal range of this assay used, might indicate subclinical hypo- or hyperthyroidism in individual patients. These findings emphasize the ability of the serum TSH to provide a very sensitive reflection of the individual's pituitary and thyroidal axis status and point out the narrow target range that most individuals require for precise L-thyroxine treatment.

The adverse effects of over-dosage or under-dosage of thyroxine are outlined here, and as they've already been alluded to, I will not dwell on them.

We performed a bioequivalency study in patients with hypothyroidism at physiologic doses because there were concerns at that point in time that there were inconsistent clinical outcomes resulting from either changes in L-thyroxine content or absorption characteristics. Our study was conducted immediately after the 1982 reformulation of Synthroid and compared typical clinical outcomes after 6-week dosing periods with either Levothroid or Synthroid in a crossover study.

Although we detected no statistically significant differences in the total thyroxine and free thyroxine index measured first thing in the morning nor any

differences in the total T3 or free thyroxine index measured in the morning, we did, however, demonstrate a statistically significant difference in the response of the pituitary to a stimulus with a thyrotropin-releasing hormone. This difference in the TRH demonstrates that there is a difference in the bioavailability being detected only at the tissue level, in this case the pituitary.

Escalante and colleagues reported in 1995 their experience with 31 patients with longstanding primary hypothyroidism considered stable on levothyroxine for at least 6 weeks prior to entering their protocol. Most of these patients were being treated with Synthroid and they were switched to a Levoxine preparation and 8 were treated with Levoxine and then switched to Synthroid. The strong point in this study is that they waited 4 months to achieve equilibrium after switching these doses before re-evaluating thyroid function tests.

This slide demonstrates the Synthroid TSH values on the left and the Levoxine TSH values on the right which is the primary illustration from the publication. What that illustration actually obscures is the fact that 6 out of 24, or 24 percent, of those that were considered euthyroid while on Synthroid were then measured as being thyrotoxic on Levoxine by suppressed TSH levels. Conversely, 2 of 21 who were considered euthyroid on

Levoxine were found to have suppressed TSH levels and therefore were considered thyrotoxic while on the Synthroid. Overall, 26 percent of these people underwent a change in their basal TSH classification, which at least would have stimulated their clinician to change their thyroid hormone dose in order to achieve a euthyroid state.

The final study that I would like to show you is the study from Dr. Dong and colleagues which was done in a more sophisticated manner than Dr. Escalante's study or even ours. Patients were recruited into this study to be euthyroid on stable doses of thyroxine at either 100 or 150 micrograms daily for at least 6 weeks prior to their randomization. Following recruitment, the patients began their assigned L-thyroxine treatment from the study drugs and after 6 weeks equilibrium, they were admitted for thyroid function testing, whereby a fasting sample prior to the last dose of the study drug was obtained and then frequent sampling was obtained over the next 24 hours. These are the four medications that were utilized.

Dr. Dong reported that the area under the curves for thyroxine and T3 were no different among the four products used in these trials. On the left are the thyroxine and free thyroxine index and on the upper right is the T3 levels. My visual assessment of the T3 data underscores the limitations of using the applied

statistical methods which are quite similar to the current standards to detect apparent differences in the profiles of this parameter.

Scrutiny of the TSH values from Dr. Dong's study, although not clearly delineated in their data set, demonstrates that these basal TSH levels along the left axis, to my visual assessment, may very well be important in light of the narrow therapeutic ranges now being suggested in that very tight target range for TSH titration. I do believe that a TSH of 2, for example, might very well be different than a TSH of 4, and certainly this degree of difference would likely be considered significant if the patient sitting in front of you was giving you symptoms consistent with hypothyroidism.

Most importantly, this graph demonstrates the individual patient TSH values from this study and they seem to indicate that a consistent TSH classification, as these various preparations were substituted, was not achieved. In this chart, the TSH colored white is the normal people with TSHs within the normal range. Those in green are those considered hyperthyroid as TSHs are below the normal range, and those in red are considered hypothyroid as their TSH was above the normal range. If these four products were indeed truly interchangeable, the color of all these blocks, of course, would be white as all of these patients

should have been euthyroid at the beginning of the study.

There is no internal control assessment here to estimate the degree of variability that would have been expected should, for example, a patient be treated with the same product from study period to study period. So, the overall variability observed here is somewhat unclear.

What I do know, however, is that all the changes in TSH classification observed here would likely have, again, resulted in clinical action by a clinician with new doses being prescribed followed by biochemical and clinical reassessment necessitating increased cost and patient inconvenience. As these results do show us, these products were not interchangeable. Clearly, we need reliable, consistently potent and absorbed thyroid hormone products in order to meet our patients' precise therapeutic needs.

Thank you.

DR. KIBBE: Thank you.

Dr. Hamilton, you're up.

DR. HAMILTON: Thank you. Thank you very much. It's a privilege to be here.

My name is Dr. Carlos Hamilton from Houston, and I regret that I do not have any support from any manufacturers of thyroid hormone to report.

(Laughter.)

DR. HAMILTON: On the other hand, I wouldn't mind having some.

(Laughter.)

DR. HAMILTON: I am currently supported by my employer, the University of Texas Health Science Center in Houston, and prior to that, my patients that I cared for, most of whom had thyroid disease.

I'm actually here representing the American Association of Clinical Endocrinologists. This is an organization representing over 4,000 physicians that specialize in the care of patients with endocrine and metabolic disorders. We're the specialists that are most often called upon by our colleagues for the care of patients with thyroid and other glandular diseases and hence we have an acute awareness of the effects of thyroid replacement medication.

We are well aware that minor changes in thyroid hormone levels in the bloodstream can result in significant symptoms on the part of our patients. When there is excessive amount of thyroid hormone in the blood, hyperthyroidism can produce a number of symptoms, including changes in the heart rhythm, accelerated osteoporosis, muscle weakness and weight loss, psychiatric symptoms and others.

When the thyroid hormone level in the blood is

insufficient and hypothyroidism results, premature ischemic heart disease can occur, high cholesterol levels, abnormal weight gain, menstrual changes, fatigue, lethargy, and other symptoms are rather common.

Dosage changes of as little as 12.5 to 25 micrograms of oral thyroxine daily can, indeed, have significant effects on serum TSH and on the symptoms that our patients describe. These changes, whether they result from change in the dose or in the brand of thyroid hormone, can have important clinical effects on our patients reducing either hyperthyroidism or hypothyroidism.

This chart or this graph demonstrates an experiment that is basically confirmed virtually every day in the offices of clinical endocrinologists; that is, minor changes in the thyroid hormone level, the thyroxine level, in the blood can result in significant changes in the TSH level. Changes of as little as 25 micrograms as shown here can produce significant elevations in the TSH when that is reduced and very low levels of TSH indicating hyperthyroidism when the level is increased.

The importance of these observations is very clear. When the dosage, the source, or the brand of the thyroid hormone replacement is changed, one should recheck the serum TSH levels in 6 to 8 weeks to verify the effectiveness of the new preparation. Changes from one

brand or manufacturer of L-thyroxine should be followed by a recheck of serum TSH to verify the equivalence of the medications. When the same dose and the same source of thyroid is used, one needs to recheck these patients only at yearly intervals.

This information is included in the American Association of Clinical Endocrinologists Medical Guidelines for the Clinical Practice for the Evaluation and Treatment of Hyperthyroidism and Hypothyroidism.

That concludes my remarks. I'd be happy to answer either now or later any questions that any of you may have.

Thank you very much.

DR. KIBBE: Thank you, Dr. Hamilton.

Our next scheduled speaker is Dr. Silva, and she is without slides.

DR. SILVA: Without slides. I'm Dr. Omega Logan Silva, a past President of the American Medical Women's Association, AMWA, an organization of 10,000 women physicians and women medical students, and as all of you know, endocrine diseases affect women to a much greater extent than men.

And I have to let you know that Abbott Laboratories is one of our corporate sponsors and Knoll Pharmaceuticals sponsored Thyroid Gland Central which was a

campaign for thyroid disease awareness.

I am a board-certified endocrinologist who practiced 29 years at the VA Hospital in Washington, D.C., most of the time as the Assistant Chief of the Endocrine Division seeing thyroid patients. Also, I served on the FDA's Immunology Panel in the 1980s and spent a number of years doing research in endocrinology at the VA after being a biochemist at NIH.

I am here to support having the FDA consider a different methodology for determining bioequivalence of hormonal products, including levothyroxine, by taking into account the endogenous levels of the hormone in test subjects.

Please read my statement since there's no time for testimony. I was told I had a minute and a half and although I talk really fast, I couldn't say everything in that minute, but if I do have a couple of more seconds, I would like to tell you a personal story.

Over a couple of weeks in the Endocrine Clinic at the VA Hospital, I had several thyroid patients come in that I had controlled perfectly on the dose of levothyroxine that I had administered, and all of a sudden, these patients were not doing well. When I checked their TSHs, they were all high, and I said, what is going on here? So, finally, I marched over to the pharmacy and

found out that the pharmacy had substituted another levothyroxine preparation without the knowledge of the endocrine service. So, I had to start all over again on these patients to get them under control.

So, it is very important for clinicians to be able to depend on the bioequivalence of these various preparations that are being looked at by the FDA. So, I would urge the FDA to do just that, to use a different methodology so that they all are equivalent.

Thank you.

DR. KIBBE: Thank you, Dr. Silva.

Dr. Brown?

DR. BROWN: Good morning. My name is Rosalind Brown, and for 23 years, I was at the University of Massachusetts Medical School, where I was Professor of Pediatrics and Director of the Pediatric Endocrine Group Division, so that unlike the speakers you have heard today, I look after the children with endocrine disorders, particularly thyroid disease, and I have just relocated to Children's Hospital Boston and Harvard Medical School where I'm now the Director of Clinical Trials Research and am developing a program in pediatric thyroidology.

My entire professional career has been devoted to the care and study of children with hormonal disorders with particular reference to children with abnormalities of

the thyroid gland. I've published numerous original articles and book chapters and have held leadership positions in both the Lawson Wilkins Pediatric Endocrine Society and the American Thyroid Association.

I'll echo Dr. Hamilton and say that unfortunately I do not have any financial relationship with any company whose product might be affected by this discussion at the present time. However, I have received research support and honoraria for speaking engagements and have been on the Thyroid Research Advisory Council, a peer-review research committee, sponsored by Knoll Pharmaceuticals in the past.

You've heard a lot about the consequences of small dose changes in thyroid hormone in adults. The purpose of my presentation is to emphasize the significant irreversible impact of small dose changes in levothyroxine on the brain development of small babies with congenital hypothyroidism.

Just to orient you a bit, congenital hypothyroidism is a disorder caused most commonly either by failure of thyroid gland development or failure of thyroid hormone synthesis. This first slide demonstrates the devastating impact of this disorder on a small infant whose congenital hypothyroidism was undiagnosed and untreated. Because at birth, affected babies have no symptoms and

because for the best outcome, treatment must be started as early as possible, screening programs for the detection of congenital hypothyroidism have been developed in the United States and throughout the world.

We now know that the incidence of congenital hypothyroidism is 1 in 3,000 babies and as such, this disorder is one of the most common treatable causes of mental retardation. In fact, congenital hypothyroidism is now known to be three to four times more common than PKU for which newborn screening programs were originally developed.

The second slide demonstrates some data prior to the advent of newborn thyroid screening, demonstrating the significant decrease in IQ of babies with congenital hypothyroidism indicated in the bottom panel as compared with the control group of normal children in the upper panel. An IQ of less than 85 is considered to be consistent with significant cognitive impairment, and as you can see, a majority of babies with congenital hypothyroidism had an IQ less than 85 indicated by the red arrow, but few of the normal babies had an IQ of 85 or less.

The third slide demonstrates the striking improvement and in fact the normalization of IQ in babies with congenital hypothyroidism indicated by the dark bars

as compared with control patients when the diagnosis was made by newborn screening and treatment was early and adequate. Unfortunately, the IQ was only normal if treatment is adequate and even small decreases in the dose of thyroxine replacement are associated with a significantly reduced prognosis.

The next slide demonstrates a study in which the IQ of babies treated with two different starting doses was compared. It could be seen that babies treated with a higher dose, 10 micrograms per kilogram per day, had a mean IQ that was 21 points higher than that of babies treated with 7 micrograms per kilogram per day, a difference that was highly significant statistically.

Similar results have been reported by numerous other investigators. For example, Rovett, et al., have noted a 4 to 5 point increase in IQ of congenital hypothyroid infants when the dose of replacement was increased by as little as 1 to 2 micrograms per kilogram per day, from 7 to 9 micrograms per kilogram per day, to 8 to 10 micrograms per kilogram per day.

These data clearly show that congenital hypothyroidism is associated with significant irreversible cognitive impairment if treatment is inadequate. Relatively small differences in the dose of thyroxine replacement can have an enormous impact and irreversible

impact, I might add, in the outcome of these babies. A potential difference of 33 percent in drug content is not acceptable for the optimal care of our patients. Bioequivalence should be determined by the serum TSH concentration, as you've already heard, which is a much more sensitive and physiologically meaningful assessment of bioequivalence than is the measure currently used to assess pharmacological equivalence.

Thank you.

DR. KIBBE: Thank you.

Our next speaker is Dr. Bryan Haugen.

DR. HAUGEN: Yes. Thank you. I'm Bryan Haugen from the University of Colorado Health Sciences Center, and I have to report that I've done past consulting with Abbott Laboratories.

What I would like to do is actually put a bit of a patient face to this by showing you one of the patients that has been seen in my clinic. A 62-year old woman presented with classic symptoms of hypothyroidism that you heard from Dr. Hamilton. She had fatigue, weight gain and constipation and her laboratory testing revealed a serum TSH that was elevated -- you can see the normal range in the brackets -- at 28 and a serum T4 that was perfectly within the normal range, which many of us see in many different patients, and we call this mild thyroid failure

or subclinical hypothyroidism.

She was treated with .1 milligram of levothyroxine once a day. Eight weeks later, she returned. Symptoms had improved, still did have fatigue, and her serum TSH was still slightly elevated, as you can see, at 7. Her serum T4 again was perfectly within the normal range and only slightly higher than her previous T4 of 8. The levothyroxine was increased by 25 micrograms, or 25 percent in this case, to 125 micrograms a day. Eight weeks later, her fatigue had somewhat improved, but now she had new insomnia, and as you can see, her TSH was now below the normal range at .08 milliunits per liter.

This is a slide you just saw from Dr. Hamilton, and I would just like to reiterate that these small changes can have dramatic effects on serum TSH as we have seen in this patient.

This also brings the point of the log linear relationship between T4 and TSH. For every linear change in the T4, either free T4 or total T4 level, there is a logarithmic change in the serum TSH, again which was illustrated by this patient, a very dramatic drop in the TSH but a minimal rise in the T4 level.

So, what are the long-term effects of this low TSH, say, on this patient with a TSH of below .1? Well, now there are many studies showing that there are ill

effects of a low TSH as well as a high TSH. Increased risk of atrial fibrillation which was found to be threefold in subjects over the age of 60 over a 10-year period, reduced exercise capacity and cardiac function, decreased bone mineral density and increased fracture risk, and again a three- to fourfold increased risk of fracture, an increased all-cause mortality in a recent study by Parle and colleagues. So, there can be significant effects even with a moderately suppressed TSH of below .1 if it is suppressed long term.

This just shows the study by Sawin and colleagues where a normal thyrotropin -- this is the risk of atrial fibrillation over time, and if someone has a low thyrotropin, which again in this study was less than .1, there is a significantly increased risk of atrial fibrillation.

So, the patient was on .125 milligrams of levothyroxine. The levothyroxine was decreased to 112 micrograms per day, a decrease of only 10 percent. Seven weeks later, she returned with no complains and her TSH now was in that target range we have talked about between .5 and 2. So, you can see that very minor adjustments in levothyroxine of even 10 percent can have dramatic effects on the target that we've been talking about, the serum TSH. So, serum TSH in patients' symptoms, not serum T4, are

therapeutic endpoints that we are using in clinical practice.

The true normal range for TSH, as was mentioned by Dr. Wood, is quite narrow at .5 to 2. Small changes in administered levothyroxine, as I've shown, 10 to 20 percent, can result in significant changes in serum TSH. An abnormal TSH, again as you have heard, has consequences. There's definitely a burden and consequences in the patient if this is not adjusted over a period of time, and there can also be a burden on the health care system by frequent testing, by utility of resources if the TSH is changing and the patient's symptoms are changing.

Thank you.

DR. KIBBE: Thank you.

I believe our next speaker is Dr. Irwin Klein.

DR. KLEIN: Yes. Good morning.

DR. KIBBE: Good morning.

DR. KLEIN: By way of introduction, I'm Dr. Irwin Klein, Professor of Medicine and Cell Biology at NYU School of Medicine, and I'm Chief of the Division of Endocrinology at North Shore University Hospital in Manhasset, New York. I'm here today as an endocrinologist and thyroidologist, and being from New York, we require support, and as such, I serve as a consultant to King Pharmaceuticals.

For the past 20 years, I've been interested in the clinical and research aspects of thyroid disease, specifically the effects of thyroid hormone on the heart. I've published over a 150 articles on the subject, including chapters on thyroid hormone in the heart, in the Thyroid Textbook, and the chapter on cardiovascular endocrinology in the upcoming edition of Brownwald's Heart Disease.

The issue that I'd like to specifically address deals with the assessment of the therapeutic efficacy of different L-thyroxine sodium preparations when used in the treatment of hypothyroidism. As you're well aware, L-thyroxine sodium is a narrow therapeutic index drug. After a diagnosis of hypothyroidism is established, treatment is initiated and the L-thyroxine replacement dose is titrated to the proper level based on a combination of both laboratory and clinical parameters. The former includes specifically the TSH level which is targeted to return to a relatively narrow normal range.

This is because, as you've heard, the effects of both under-treatment and over-treatment are potentially harmful. Specifically, excess T4 replacement producing a low serum TSH, as reported by Sawin in the New England Journal of Medicine in 1994 and as reviewed by us in that journal in February 2001, can produce atrial fibrillation

in as many as 30 percent of patients above the age of 60.

My review of the FDA guidance of bioequivalence of L-thyroxine sodium indicates that it is possible to consider two preparations bioequivalent, based upon T4 pharmacokinetics which fall between minus 80 to plus 125 percent of the reference compound.

As a physician who cares for many patients with hypothyroidism, I am concerned that the application of the existing guidelines for bioequivalence will yield results which do not properly reflect therapeutic equivalence. It has been well documented that even with a normal blood level of T4, a low TSH level predicts increased cardiovascular risk. This opinion then can demonstrate that any study of bioequivalence must include serum TSH measured at steady state.

We have provided a review to the committee which I believe further outlines the basis for this conclusion. If, however, the existing guidelines are not amended to reflect the principles which I've discussed, the resulting effect may be that substitution of non-therapeutically equivalent L-thyroxine preparations will produce unwanted effects among the over 10 million patients currently treated for hypothyroidism in the United States.

Switching a patient from one formulation of L-thyroxine sodium to another approved under the current

guidelines would require that the physician perform repeat TSH testing and dosage adjustments to assure that these patients remain euthyroid. Otherwise, it could well be expected that as many as 20 percent of these substituted patients would experience a fall in TSH. For the over 60-year-old segment of the population, that change would place 10,000 patients each year at risk for iatrogenic atrial fibrillation.

Since the cost of treatment of each of these patients is conservatively estimated at $7,000, the increased health care costs beyond the cost in human health as a result of these actions could well be in excess of $70 million annually.

I'd be happy to discuss these opinions with you further. Thank you.

DR. KIBBE: Dr. Tuttle.

DR. TUTTLE: Thank you very much. I'm Mike Tuttle. I'm one of the endocrinologists from Memorial Sloan Kettering Cancer Center. Unlike most endocrinologists, I see a very skewed view of the world working at a cancer center. On any given month, 80 to 90 percent of my patients have thyroid cancer and at least half of them have metastatic disease. My clinic is a great place to come learn to do thyroid cancer. We're not a great place to talk about diabetes.

I also need to let you know that I have received Knoll grants in the past before and do a lot of lecturing and speaking about thyroid cancer around the country and have received honoraria for that.

Typically when people think about thyroid cancer, it's frequently thought of as one of those really unusual cancers you never see, but if you look at the actual number of cases, the number of new cases being 22,000 isn't that much different from other more, I suppose, popular cancers, multiple myeloma, kidney cancers, leukemia and lymphoma. 1,400 deaths this year are expected from thyroid cancer. Fortunately, the overall survival in thyroid cancer is 90 percent which means the vast majority of patients with thyroid cancer will be long-term survivors and will require levothyroxine therapy.

Now, I'm a clinician, and to me, what matters is how we take care of patients. Initially in thyroid cancer, we usually start with a total thyroidectomy, surgically removing the entire thyroid. We use radioactive iodine as a very targeted therapy to destroy any residual normal tissue or any metastatic thyroid cancer and that functionally leaves the patient with no thyroid tissue. That is the goal of our therapy.

Now, if you think about that at first blush, you'd think the real role for levothyroxine is what you've

been hearing this morning, which is just to replace that patient, get rid of the hypothyroid systems and keep them normal, but in fact in thyroid cancer, levothyroxine therapy goes far beyond that.

Numerous studies over the last 30 years have shown that if we use what we call levothyroxine suppression, in fact an overdose of levothyroxine, to suppress that TSH, we see a marked decrease in recurrence and better outcomes. So, the goal in thyroid cancer is, A, yes, to replace them, so they don't have the hypothyroid symptoms, but more importantly for us, I frequently call this to my patients, this is our chemotherapy that they're going to be on for the next 20, 30, 40 years, depending on how old they are.

If you put this into some perspective, you've heard this morning, our usual goal for primary hypothyroidism is a TSH around 1, a T4 in the normal range. In my clinic, our goal is much different. Our goal is to have a TSH that's very, very low, bordering on undetectable, and to do that, we have to get their T4 elevated. On purpose in my clinic, we make folks subclinically hyperthyroid. The goal is to get them on just enough T4 so that they don't feel it clinically but yet we produce the biochemical suppression we want.

What that means is very small changes in their

dose, as little as missing one thyroid pill a week or taking one extra thyroid pill a week, can tip them over the edge into clinical thyrotoxicosis. This is not just numbers on a piece of paper. This is phone calls to my office from real patients having rapid heart beats and nervousness and not being able to sleep. Alternatively, if the dose is decreased a little bit, they feel perfectly fine, but the TSH is now up into the normal range and they're at risk for recurrence.

Now, to try to put this into some perspective, how big a dose change do you need? You've already heard this morning that small dose changes, which is my usual dose increments, of 10 or 12 percent are enough to produce these symptoms, either for the worse, which is thyrotoxic symptoms, or back into the normal range. Unlike most of the TSH measurements you do in hypothyroid patients which may be once a year, in thyroid cancer patients, we maybe do these every four to six months because fine-tuning is critical.

So, what I hope to leave you with today is that the goals in levothyroxine suppression in thyroid cancer are much different. This is chemotherapy for us. The implications of having a TSH a little out of the normal range is far more significant in thyroid cancer. The narrow therapeutic window that you already use for thyroid

hormone is much smaller when we're dealing with folks with thyroid cancer. These very small changes can have important clinical events. These are not just paper changes that we chase. These are real events in the lives of our patients, and to our mind, product substitution with alternates that vary by really more than 5 to 10 percent would be unacceptable in the treatment of thyroid cancer patients.

Thank you.

DR. KIBBE: Thank you, Dr. Tuttle.

Dr. Dickey. I hope we're in the right order. Richard Dickey?

DR. DICKEY: Yes, sir. Thank you. Good afternoon and thank you for inviting us to testify today.

My name is Richard Dickey, and I'm a newly retired physician. I practiced endocrinology for over 30 years and still practice as a volunteer in a local indigent clinic in Hickory, North Carolina. I also continue to teach at Wake Forest University School of Medicine.

I'm pleased to testify before you today on behalf of the Endocrine Society, where I serve on the Clinical Affairs Committee. The Endocrine Society, founded in 1916, consists of over 11,000 physicians and scientists dedicated to research and patient care in the field of endocrinology. Our clinician members are involved in the

daily treatment of patients with hormone disorders, including thyroid disease. We publish four peer-reviewed journals, Endocrinology, Endocrine Reviews, the Journal of Clinical Endocrinology and Metabolism, and Molecular Endocrinology.

I have no current affiliation, financial or other, with any manufacturer of levothyroxine products. The Endocrine Society receives financial support in the form of unrestricted educational grants from several manufacturers of thyroid drugs, including Abbott, King, and Watson.

It is our dedication to the treatment of patients with thyroid disorders that brings us to this hearing today. In the interest of time, I'll not go into the manner by which the FDA tests for bioequivalence, as you've heard from leading thyroid experts today on that matter. Instead, I'll focus our comments on the issue of direct patient care, as have many others today.

Testing for bioequivalence is important and we support the FDA in their diligence in this matter. However, when testing hormone-based drugs, bioequivalence data needs to be supplemented by therapeutic or clinical data. Bioequivalence does not equal therapeutic equivalence. Bioequivalence testing does not currently include a mechanism for factoring in a baseline correction

for endogenous hormone production in the patients tested and therefore therapeutic differences can be missed. These differences are clinically significant when treating patients with thyroid disorders, such as thyroid cancer and hypothyroidism.

Endocrinologists are trained and experienced in caring for patients with complicated thyroid disorders and, regardless of bioequivalence data, realize that levothyroxine products are not interchangeable. Our concern is that without any supplemental information, other physicians without the same level of specialty training in endocrinology may assume that bioequivalence does equal therapeutic equivalence. In the patient, the consequences of important differences in bioequivalence and therapeutic equivalence between products become obvious over time, as demonstrated in the health or ill health of the patient. The differences can even result in serious complications, complications that could have been avoided.

We urge you to focus on patient effects and accept that bioequivalence is not therapeutic clinical equivalence for a hormone such as levothyroxine.

In conclusion, I would like to again point out that our participation today was in the interest of the patient. For your information, a disclosure statement regarding those clinicians involved in the review of this

issue and the development of this testimony, as well as financial relationships to the manufacturers of thyroid products, is included in our written testimony provided to each of you.

Thank you.

DR. KIBBE: Thank you, Dr. Dickey.

Dr. Bolton?

DR. BOLTON: I guess I have overheads.

First of all, I guess I should tell you some disclosures. I'm speaking here on behalf of Geneva Pharmaceutical Company who has recently developed a thyroid product and gone through some bioequivalence tests. This is the very first time, by the way, I've ever really worked with Geneva. I must disclose, also, that I own stock in Abbott Laboratories and Forest Laboratories, and so that might sort of neutralize some of what I'm going to say.

(Laughter.)

DR. BOLTON: First, I'd like to tell you what I aim to do here and that is, I aim to show you, in what I consider a very objective and scientific way, a look at the data that's been shown to me by Geneva, and I'd like to defend these studies as demonstrating that these products are equivalent and there's a very consistent measure of performance.

First, let's look at the design of these

studies and understand that the FDA recognized that there was a great variability in thyroid products. I think we all know that and in recent years have come upon recommending a guidance so that we can overcome some of this variability and put some regulation on the production and design of thyroid products.

So, the recommended protocol now for a bioequivalence study is the standard study, but I'd like to point out a couple of things here.

Number one is the sample size, 24. When you do a 24-subject bioequivalence study, you're suggesting that you have a relatively low level of variability, which we'll see in the data is true.

The other thing I'd like to point out is the dose, a 600-microgram dose. That's a large dose, but because of analytical problems, it's very difficult to do these studies with smaller doses, and we'll talk about that as we go along. So, what we do here is give multiple tablets of lower doses to equal the 600 micrograms.

The other thing that is a little different about this is the baseline correction. That's been brought up before. Now, they're asking not only for the total T4 but they're asking for baseline subtracted data and then performing a statistical analysis using covariants, and the requirement, as far as I know, is that all three of those

methods must result in passing the bioequivalence criteria. So, it puts the onus on this product a little more than it would on a usual product.

Also, we understand that the acceptance is based on a confidence interval, not on a statistical test, hypothesis test. The other thing is that we're using subjects and not patients. That's been mentioned before, and I think that's been really bandied about a lot by the FDA and the experts and so on, and we know that subjects are just a way of measuring whether two products are equivalent or not. It's a mechanism or a machine that we put the product into and we look at the output. We're not looking to see whether it's different between normal subjects and patients but just whether the formulations are performing the same way. I think we all understand that.

So, let's go to the next slide. You see, my understanding of bioequivalence is that if we have two products where the blood levels are absolutely identical, that any pharmacodynamic or therapeutic effects will be identical and any secondary effects will be identical because if the blood levels are identical, it's very hard to think that therapeutic effects will be different. In my experience, I have known no examples that belie this particular assumption for oral products, particularly.

If we don't believe this and we don't go by

this assumption, we would have to do clinical studies for most drugs or at least we can make an argument for most drugs, and from my point of view, that would be sort of going against the concept of bioequivalence which is using a bioequivalence study as a surrogate for a clinical study for approval of generic drugs.

I'm going to go through some of the studies that I've seen and give you an idea of the results.

The first study was a dose proportionality study. First of all, the dose formulations are dose proportional. They're the same formulation, just larger tablets as the dose goes up. The pharmacokinetics show very good dose proportionality, and I think in the next slide, you're going to see the results of the dose proportionality study.

These are three different doses just made up to 600 micrograms, and I think it was 50, 100 and 300, and they're virtually superimposable. You might say, well, this is just the average results. By the way, the averages were -- if you would look at the ratios there, they're just about a 100 percent exactly, and you might say, what about variability? The variability here was very small. In fact, for the total T4, I think the variability was around 10 percent CV which is a really low variability drug, which is very good because we have a narrow therapeutic index

drug.

Next slide, please. This is a study of the generic or the new preparation prepared by Geneva versus Synthroid. This was the result of the typical study. The top slide gives you the average results for total T4 and the bottom slide is the corrected T4, and I can tell you for the total T4, the CV was less than 10 percent. I'm going to show you more about that in just a moment.

Next slide, please. Here's another study done against Levoxyl. Again, this is just a head-to-head study, typical bioequivalence study, virtually superimposable average blood levels. The ratio of Cmax and AUC again for this was very close to 100 percent, like 101, 102, something like that, very low variability.

Next slide, please. Here, I'm just going to give you an idea, a little bit of the averages and the variability. Interestingly, the variability was lower when we just used the total T4. In fact, in all studies that I saw there using total T4, the variability was on the order of 10 percent, sometimes a little less, sometimes a little more, but the averages were always very close to 100 percent, and these products are very similar. The dissolution for these products are almost 100 percent within 30 minutes. So, we have a relatively simple formulation. There's nothing complicated about this

formulation, very rapidly dissolving, and we wouldn't expect to see a lot of variability.

Next slide, please. I just did a little simulation or computation to see what we would expect if we tried to do these studies on lower doses lower than 600 micrograms. If we are subtracting the baseline and the CV, the coefficient of variation, the variability is due only to the assay, this is the kind of variability that I would expect to see with a 600, 300 and 150 microgram dose because the subtraction of the baseline reduces the values that we see, and if we tried to do, for instance, a 150 microgram study, the variability just due to the assay -- that's the assay of the active material, nothing to do with biological variation -- would be at least 44 percent.

Now, there is one slide missing that I unfortunately did not put up here, but it had to do with the ratios in these studies. You know, it was the old 75-75 rule, which I don't mean to impose on this, but I'd like to point out that 80 to 90 percent of the patients, subjects rather, 24 in each of the studies, had ratios that were between 75 and a 125 percent. Most of them were between 80 and 120 percent. That's individual ratios and somebody can say, well, 80 percent, that's 20 percent off, but when you see 80 to 120 percent, that's including the variability of the assay, the biological variability. So,

if we see individual ratios between 80 to 120 percent, we have a terrific product and that's what I saw for this product.

Thank you.

DR. KIBBE: Thank you, Sanford.

Our last speaker, Dr. Bill Barr.

DR. BARR: Good morning. I'll try to be brief. I know everybody is hungry.

Like some of the other speakers -- by the way, my name is Bill Barr. I'm Director of the Center for Drug Studies at the Virginia Commonwealth University, and as such, I receive money from almost everybody.

(Laughter.)

DR. BARR: I have received money specifically from MOVA, from Abbott, Vintage, and Alara, all of whom make these products, make levothyroxine products, but would like to emphasize that my views today are my own and haven't been either approved or sanctioned or disapproved by anybody.

I'd like to present some data that I think are relevant to the issues today and then present some views which I hope will be useful.

This is a study which we ran several years ago and which I'm going to refer to just as test and reference in which we studied two levothyroxine products, a test and

a reference product, that were tested in patients that had been stabilized previously on 100 micrograms of levothyroxine. We then switched them over. They either started them with test or reference, and then we switched them over after a month, after they reached steady state again.

During this procedure, we did in fact measure TSH. We did it actually for safety reasons, but we did measure TSH, and when we looked at TSH, we did find something that was very interesting. If you look, you see that when we shifted them over, there are some patients that jumped way up in TSH values, that when we shifted over to the reference product, TSH levels in some patients went up quite considerably, and went up in fact above the range in which most clinicians would have begun to question that particular product or that particular result to the point where they may have switched them and actually had to do dose adjustment because the TSH levels at that point were above the 4 to 5 to 6 that most clinicians consider to be relevant whenever they're making dose adjustments.

Now, I thought this was very interesting whenever we looked at this. I wanted to see if there were any other products that were done similar that were tested in a similar way and found another study.

May I have the next slide, please? This was a

study that I found in actually the Virginia Formulary through FOI and it was done by Forest Laboratories. This again was a product in which both products were given to patients and it was done at steady state in which they also measured TSH levels. I apologize for the quality of these slides. But you can see this particular product, the old product, was all below this level.

May I have the next slide, please? However, with the reference product again, Synthroid, many of the levels did go well up, not all but a few. What we've seen in both studies is a subset. There appears to be a subset of individuals who take the reference product, in this case Synthroid, who uniformly jump up with the TSH levels and that may be part of the explanation that many of the clinicians have talked about today.

May I have the next slide, please? Let me give you a possible explanation for this subset. This is my hypothesis. These are the in vitro dissolution times for the reference product. This was the older Synthroid product. I can't say whether this is relative to today's product or not, but I simply want to give you an example of why these TSH levels changed.

About 50 percent of the drug is not dissolved in this in vitro method at about one hour. On the other hand, the other products that I've just talked to you

about, almost all of them follow the current USP dissolution definitions, which means that about 80 to 90 percent or 100 percent have to be dissolved by, I think, 20 minutes or something like that. In fact, almost all of the generic drugs that are made today are made by a dry granulation in which almost 80 to 100 percent are dissolved within 20 minutes. Now, this is not true, unfortunately, of the reference product. The reference product, you can see, is much more slowly dissolved.

May I have the next slide, please? By the way, levothyroxine is not absorbed in the colon. It's absorbed only in the small intestine. It's one of these drugs that we consider to be transit time dependent. So, if it's transit time dependent, if you look -- these are some data by Davis that are transit times of the small intestine. All of these dots represent each individual person in all the studies compiled. And this is one hour, and you can see that there's only about probably 5 to 10 percent of the people at any give time that have transit times in this particular study of an hour. It depends on how you do transit times, by the way. But in this particular study, the transit times were only about an hour.

Therefore, we would expect that with transit times of about an hour and when only 50 percent of the drug may be dissolved in an hour, that there would be a subset

that would probably have TSH levels at some point in time, depending upon their transit times.

Now, transit times is a highly variable situation. For example, if you have a drug that is going to fall over into this area, it would seem to me that you're going to have greater variability in this drug as well, and this greater variability could be seen, for example, in the next slide.

Women. There are several studies to show that the transit time in women vary within the menses, that the follicular state may be different than other parts. In fact, this is one study. There are some controversies about this data because they were used with lactose which is not the best way to measure transit time, but it does illustrate the example. This is at the follicular phase and this is at the luteal phase at the transit times, almost double the transit times. So, transit time may be a factor.

The point that I do want to point out is that there is a subset for whatever reason and it probably is related more to dissolution rates. It is my guess that we probably don't need a lot more complicated studies. I think that in fact you could probably do much simpler studies if all of the products, in fact, had dissolution standards in which everything was dissolved within 20

minutes. The transit time would not be a problem.

What I think that we will see is that as long as we have two sets of standards -- and at one time, the USP proposed that they were going to have two sets of standards, one for one set of compounds and one for another set of compounds -- if that's true, we will always have problems of interchange. I believe that whenever you look at today's market, which unfortunately, good or bad -- and I'm not sure it's good -- allows widespread interchange, that this will be a continuing problem. I think that probably we need to address the problem in a more complete way and look at all of the factors that may be involved, including transit times, including dissolution.

Thank you.

DR. KIBBE: Thank you, Bill.

Well, that brings our open hearing to a conclusion. We're only 10 minutes late. I did make a statistical analysis and the M.D.s took 4.3 minutes to do their presentations and the Ph.D.s took 10.8 and I think there's a correlation in there somewhere.

(Laughter.)

DR. KIBBE: But let me assure everyone who came that we do not take this situation lightly. We will take into account all of the information that was presented to us and supplement it with additional information that we

can get from valid scientific sources and certainly it will be a high priority item for our Biopharmaceutics Subcommittee to look at. We really do appreciate your interest and your efforts on behalf of the American public.

And I think we now stand adjourned for lunch. We're going to open up again at 1:30 with bioequivalency and continue the discussion on endogenous drug substances.

(Whereupon, at 12:40 p.m., the committee was recessed, to reconvene at 1:30 p.m., this same day.)

AFTERNOON SESSION

(1:30 p.m.)

DR. KIBBE: If I could call us all back to order. I've got everybody back, and I know I see over in the corner that our first speaker is here. So if I could call us all back to the meeting and ask Dale to kick off our discussion of bioequivalency with endogenous drugs.

Thank you, Dale.

Excuse me. Efraim?

DR. SHEK: I want to just note for the record that since my employer has an interest in this discussion, I am recusing myself from active participation in this session. But with your permission, I'll continue sitting here because it's a packed house.

DR. KIBBE: Thank you, Efraim.

Dale?

DR. CONNER: I'm sure you're all getting tired of seeing my face, especially going on and on about trying to tell people the basics of bioequivalence which I'm starting, I think, after these many years, to get tired of trying to explain to people and still hearing a lot of misconceptions about it.

I'd like to start off, though, on the part of the FDA by saying another vote of thanks to the people that came during the public comment period. I know that they

took time out from their busy schedules, sometimes at a lot of expense to themselves to come and give their opinions and concerns, and I'd like to say that we at the FDA take those concerns very seriously and they're of great value to us. And so thank you again, if any of you are still here, that you actually came and gave us your input on that.

The topic today that I'm starting off with is a much more general topic than was discussed during the comment period in that it's the bioavailability and bioequivalence of endogenous substance drug products in general and what are the concepts behind generally looking at those things in endogenous drug substances.

So I'm again the lead-off person for this topic. You'll be seeing later on a couple of very nice examples of this that we've had some experience with, and we're going to try and work this into a discussion of what are the general principles of dealing with these type of products and what are the variables and things you have to look at in deciding how to determine bioavailability and bioequivalence. So this is again, to use Ajaz's previous term, an awareness topic discussion or it's the first step in the discussion that may follow on this general topic, and the purpose of this whole discussion is to provide information to the committee on the challenges for BA and BE assessment of endogenous drugs in general.

Perhaps at later times, we'll take this, after this initial discussion and information sharing, to the Biopharmaceutic Subcommittee meetings or to perhaps another ACPS meeting where we can talk about and debate in general in a more in-depth fashion. So at this meeting, we seek your recommendations on how to develop this information needed to enhance the science in this area.

So as you may have figured out already from some of the comments, the bioavailability and bioequivalence of endogenous drug substances needs special considerations. And I'll go over my infamous diagrammatic explanations in a second. These considerations were not addressed in our general bioavailability/bioequivalence guidance, and if you're familiar with that document, which I think we're very proud of, it still left out those considerations for those type of products and hence our need to really discuss what we've done so far successfully on several of the products and how that success can be extended to other products where it's not quite as clear-cut.

The specific things that we do have guidances on that relate to this topic are specifically two compounds or two endogenous substances, the first being a bioequivalence guidance on potassium chloride modified release tablets and capsules and that's listed up on my

slide. I have to say that the second one for levothyroxine sodium tablets refers only to the bioavailability of those products. It does not address the bioequivalence. There seems to be some confusion amongst a variety of industry people, as well as some of the public comment people, that that in some way was supposed to describe bioequivalence policy for levothyroxine. That's not the case. It's strictly a bioavailability guidance, as stated in the title.

Just a short list of some products that might be considered as endogenous substances which may involve special problems in doing bioavailability and bioequivalence. Estrogens, for example, testosterone, progesterone, calcitriol, and someone suggested to me that -- I wasn't even aware of this. Someone who had worked on the NDA said ursidiol. Also, some other products which are not given orally but are given as parenteral non-solution products, such as insulin and human growth hormone, could be said to have some of the same considerations.

Again, the next slide or two or three is something that the committee saw yesterday in my other talk. It's just important to point out that these are pharmaceutical equivalents. So we're not dealing with therapeutic substitution or any substitution of different types of dosage forms. When we do these comparisons or

bioequivalence comparisons, we're dealing with the pharmaceutical equivalents containing the exact same amount of drug substance in the same type of dosage form.

And I think that I went over this particular slide, that we're really in the long run or at the end, we're interested in assuring therapeutic equivalence, and we, through our very extensive experience in a wide variety of drugs, some endogenous, some others, we've arrived at, through many years of experience in assuring TE, or therapeutic equivalence, the most efficient ways to do proper bioequivalence tests with proper analysis and acceptance criteria.

I said yesterday this is my favorite slide and I can't be restrained from throwing it into every talk. It actually is relevant, and I have three versions of this. Here's my general. I don't want to call it generic version because I work for generic drugs, but this is the simple version for the usual non-endogenous oral drug product. It simply flows again from this first step where we have a solid oral dosage form and that dosage form, I think we can all agree, needs to release the drug and make it available to the body, and so it seems like a simple concept but the drug has to leave the formulation and get into the body to eventually create a therapeutic effect. And sometimes by therapeutic effects, I mean any effects that a drug caused,

both desirable and undesirable.

So the first step usually for an oral product is that product has to disintegrate and then go into solution and once in solution pass across the gut wall.

So when you look at bioequivalence specifically, what you're really looking at -- and it's an important concept that people get confused about -- is you're looking at formulation performance and some way to adequately assess how these comparator formulations behave when taken by patients, or if you're doing a study by normal subjects, how they behave, and can a formulator make another product that behaves in exactly the same way. So that's the whole point of bioequivalence testing, and if you keep repeating to yourself it's all about the formulation and whether that formulation performs in an identical or close to identical fashion and releases the given drug in the same manner, same rate, and same extent.

So how do we infer, how do we measure whether that's actually happening? Through my process here, we go through drug passage through the gut wall. There are plenty of other steps that you could put into this. I've kind of over-simplified it. It passes into the blood. The blood acts as an intermediate transport area, carries it to the site of activity, and one gets therapeutic or pharmacodynamic effects.

Then as I mentioned yesterday, we've chosen, I think, as a matter of efficiency to do blood concentrations, when we can, for bioequivalence purposes simply because they are very close to the event we're trying to measure which is the only thing we really have control over which is the formulation. All the rest of these things are patient or subject physiology-related events. The thing that we really have control over is what does the formulation do, and formulation scientists can design it with various properties, release slower, release fast, or so forth, and so this is the both the thing that we're trying to measure and the thing that we actually have control over.

So we've chosen to measure in blood for several reasons. Blood is not too far removed from the event that we're trying to measure. It's also related in almost all cases to the therapeutic effects that are eventually achieved by the drug since the blood is thought to be an equilibrium or related to the drug appearance at the site of activity. So in all respects, the blood answers most people's questions very adequately and very efficiently.

It also happens that blood levels for regular drugs, not endogenous substances, have some very nice properties. I mean, either it's a straight line relationship between what you're trying to measure and the

dose or at worst, it's a nonlinear function where, on this particular graph, a nonlinear elimination would make the curve go upwards which actually increases the sensitivity of the test. And by sensitivity in this respect, I'm saying that a test done in a nonlinear range is much more likely to fail the product. So it becomes extremely sensitive to small differences. So in effect, even a nonlinear drug tends to make products fail rather than passing products that are quite different.

The therapeutic or pharmacodynamic effects have different properties. Any clinical effect, just about any clinical effect tends to be more variable because, as you proceed along this scheme of mine, you pick up variability with each step, and so the clinical effects or clinical measures that we usually use -- and I think you saw some of those described yesterday in one of the talks -- tend to be quite variable, and they also have different properties in the blood.

Generally with pharmacodynamic or clinical effects, if we remember from our pharmacology textbooks, you usually have an S-shaped dose-response curve. So you have essentially three parts of that curve. You have the part where you're really not giving enough to cause an effect, so you get close to no effect. You have a steep portion in which you can actually see very large changes in

your clinical response with very small changes in dose, and I think you saw some of that described in the public comment period. And then at higher doses, you have a plateau where you've gotten the maximum effect. You really can't get anymore. If you're testing for equivalence or testing to products up at the top of the range, you really have no sensitivity or no ability to tell the difference between them simply because when you're on the plateau with a maximal response, you really can have tens or hundreds of times difference in the bioavailability and not see any difference in the response.

So it's critical, if you're going to use this type of response to test the difference between formulations, that you do it at the proper dosing range where you're on the steep, sensitive part of the curve. So that's one of the considerations for doing equivalence testing between products using a pharmacodynamic or clinical response.

How does this situation change? I mean, it seemed a fairly simple, straightforward, beginning to end process, but how have I changed that to look at endogenous drug substances, such as hormones?

Obviously we have now a substance that -- if we try and measure it in blood. In the previous drugs I described, the only source of that drug appearing in blood

is from the dosage form that you actually gave. Now, it's not quite so simple. We have not only that dosage form that we gave supplying drug that appears in the blood and throughout the body, but we have the body actually producing that drug. So we have at least two sources or more sources for that substance to appear in blood.

And to make things even more complicated, especially with hormones, there's also a feedback process where it isn't simply a steady body production, that as blood concentrations go up and down, that production and that storage of that compound changes with changes in the blood concentrations or the body concentrations. So that adds a level of complexity that really creates certainly technical problems in using our normal methods for doing bioequivalence, and certainly that process and the amount in blood that did not come from our formulation has to be taken into account if one hopes to use pharmacokinetic measures to determine bioequivalence and determine difference between formulations.

So I've redrawn this and it's drawn for illustration, not entirely supposed to be accurate or representative of any given product, but I've changed the supposedly nice properties of pharmacokinetic data to say, well, now we're dealing with a baseline or that substance is already there before we start to add the contribution of

the dosage form on top of that.

Well, that's not the only case. Our other example that I mentioned is potassium chloride, and how does potassium chloride differ from, say, hormones of the system I just described? With potassium, on the other hand, the body actually, strictly speaking, doesn't make potassium. So it more or less shifts it around. It takes it in from the diet. It puts it out in the urine and perhaps the feces, and so you're really looking at an equilibrium process where, if a patient is deficient in potassium and is given supplemental potassium, they tend to take more in and store it, hopefully. But if you deal with normal volunteers with proper and healthy levels of potassium, most of what's taken in is simply put back out again. So the body doesn't really need to hold onto it or to increase stores. It basically comes in one end and goes out the other, so to speak.

So the question is, what we do with potassium, on the other hand. Again, we're dealing with the same set of issues in a way in that there's a lot of potassium already in the blood. If we give a single dose of potassium, you really don't see that much of a change in the blood. It's a very, very small change. So even if you were to correctly subtract the baseline, the signal you would end up with is extremely tiny. In effect, probably

in the upper 90 percent of the area of a given dose would have to be subtracted which would leave you with a very small signal, very highly variable, very difficult to do studies on. Probably any kind of reasonable size pharmacokinetic study done on the blood would probably fail even on a product against itself.

So the blood has proven to be not a very good site for sampling of this. It's good for most products and most types of drugs. However, in this particular one, urine has proven to be a much more effective means of assessing bioequivalence because, as I said, most, if not all, of the potassium you give in the dosage form to a normal healthy person comes out in the urine.

However, it's not quite that simple because that's not the only source of potassium that comes out in the urine. You actually, especially with normal subjects, have to eat, and if you have a several-day study and you try not to feed them, they get very angry and cranky. So you really have another source of potassium during your studies that comes from the diet.

So the urinary data that we collect also has to be adjusted for baseline and that baseline potassium that it has to be corrected for is basically what you gave in the food during the study. So you still are facing baseline correction in the urinary data for potassium as

well, and as I drew it here, although it's definitely not to scale, if you look at the blood concentrations, you're dealing with a much, much higher baseline than my previous illustration and that makes the blood more or less unsuitable for this particular bioequivalence procedure.

Again, I was going to just like pass over this slide quickly, but I again notice some people who didn't seem to understand the criteria that we used for bioequivalence, especially this last one, 90 percent confidence intervals must fit between 80 and 125. There's a given misconception in the community that bioequivalence of 80 to 125 allows the mean data of a comparison between two products to vary between 80 and 125 percent. That's absolutely not true. That's a misunderstanding of the criteria.

What we're dealing with is the confidence intervals around that data, and that's based on the variability of the products and the variability of our study. Generally, for most products with normal levels of variability, say CVs of 25 percent or as much as 30 percent, the mean data or the point estimates that we see in normal bioequivalence studies don't generally fall outside of 10 percent and most of them are around 3 percent either way because essentially the confidence interval has a width around that mean and it doesn't really take much

movement away from center to cause the edge of that confidence interval to go over our limit and fail. So if you're really just talking about mean data, the means never really get a chance to get out anywhere close to the plus or minus 20 percent.

So the problems that we deal with or the issues, among others, are assay sensitivity which has been mentioned before, that if you do your study and you don't give the assay a high enough signal, then you have some problems with variability and inability to tell the difference between two products. That's one of the reasons, say, for example, with levothyroxine that the original recommendations were for 600 micrograms. So lower than that, based on the data that we had, we really did not think that anyone could really see the difference between formulations at a lower dose simply because of lack of sensitivity of the assays to even detect that in the blood.

Obviously, endogenous baselines are always a problem. You need to be able to deal with correcting for the baseline if necessary or deciding whether baseline correction is necessary.

The feedback inhibition or feedback control of the endogenous production is an important concept which relates to the baseline still.

Some of these under normal conditions have

circadian or other types of rhythms or variability throughout the day and that has to be taken into account.

And some of these are claimed to be either linear or nonlinear pharmacokinetics which, as I said, is another consideration that controls the sensitivity of the test.

So today, as far as the agenda goes, we will have two case studies, the first being a case study on levothyroxine with actually two speakers in that case study. The first is our speakers from Abbott Laboratories who will go over a very interesting study that they did on baseline correction and some other issues. It's an extremely interesting study. Steve Johnson will then speak for the FDA about our experience with levothyroxine bioavailability in quite a few NDAs that we've reviewed now.

The second case study is on potassium chloride and more detail will be gone into on our experience with potassium chloride, and finally I'll come back and just kind of wrap things up with a summary.

First off, Steve will introduce the topic of levothyroxine.

DR. JOHNSON: Good afternoon, ladies and gentlemen, members of the advisory committee. My name is Steven Johnson, and I'm a clinical pharmacology and

biopharmaceutics reviewer, collocated with the Division of Metabolic and Endocrine Drug Products.

Today I'll be presenting on a very important endogenous drug substance that you've heard a lot about this morning, and this product has come to a focal point here at the Food and Drug Administration within the last several years.

My presentation this afternoon will cover two primary topics. The first is a background or a description of why levothyroxine sodium was declared a new drug in 1997. I'll discuss specific aspects of the guidance for industry for this product. The second part of the presentation will focus on the FDA's current recommendation for evaluating bioequivalence between these levothyroxine products, and at that time, when I discuss that section, I'll talk about the recommended study design and on the bioequivalence analysis itself.

Well, prior to August of 2000, levothyroxine sodium was an unapproved marketed drug. It had actually been grandfathered in. It was introduced in the 1950s as a more pure synthetic form of thyroid, USP, and in 1997, it was estimated that there were at least 37 manufacturers or repackagers of levothyroxine sodium tablets.

However, despite the fact that we had more than 40 years of clinical experience with this particular

product, there was still a high degree of uncertainty about the products themselves and the uncertainty existed with all of the products that were currently on the market. Namely, there were issues about product stability, which has a direct impact on the shelf life or the expiration dating of the product, formulation consistency and content uniformity concerns within a given brand, and then there was the issue of bioequivalence. Bioequivalence had never been formally established between brands.

Well, levothyroxine sodium degrades very quickly when it's exposed to light, moisture and oxygen, and when it's combined with a carbohydrate excipient, it undergoes a biphasic degradation process whereby there's a rapid initial decay phase followed by a more gradual degradation phase.

These characteristics have a direct or a negative impact actually on the product's stability. Between 1990 and 1997, there were 10 recalls involving a 150 lots and over 100 million tablets. These recalls ranged from Class 1 to Class 3 and were initiated because of content uniformity, subpotency, and stability failures.

In an attempt to address these issues or these stability problems, many products were manufactured with a stability overage which is very distinct or different than a manufacturing overage. It's a very important distinction

because a stability overage is intended to extend the shelf life of the product and we saw a lot of that and that's not acceptable to the agency, whereas a manufacturing overage is sometimes necessary to account for some of the loss during the manufacturing process itself.

In 1987, Fish described overages in levothyroxine products as high as 9 percent. The FDA actually has internal documentation that would suggest that in some cases, these stability overages were actually as high as 15 percent.

The FDA also has evidence that significant changes were being made to the product formulations in an attempt to improve product stability, and these changes were to both the amounts of the active drug and also to the amounts of the product components.

There was also evidence from case reports in the literature that suggested that therapeutic failures had occurred when patients had received a refill of the same product for which they had been previously stable. Of the 58 cases of therapeutic failure reported to the FDA between 1987 and 1994, nearly half had occurred when patients had received a refill of a product on which they had been stable for years.

So in 1997, in an effort to standardize levothyroxine sodium tablets and to reduce the instances of

therapeutic failures, the FDA declared levothyroxine sodium tablets a new drug and sponsors wishing to continue to market their particular product needed to submit either an NDA or file a citizen's petition describing why an NDA was not necessary for their product.

At about this same time, essentially in concert with the Federal Register Notice, the FDA recognized, in part due to the large number of manufacturers of this product, that we needed to come up with a consistent set of guidelines for this product and so a guidance for industry was put together. This guidance was intended to address issues of bioavailability, as Dr. Conner pointed out earlier, and was never intended to be used on its own for the purposes of bioequivalence.

I've chosen three topics here, I've highlighted them in red, to discuss a little bit further from this guidance. The first of the two bioavailability studies evaluated the in vivo performance against an oral solution. Two 300 microgram tablets, the test product, were compared to a 600 microgram oral solution in a single dose to a crossover study design. Pharmacokinetic parameters, AUC and Cmax, were evaluated without an endogenous baseline correction, and total thyroxine was used as the measure.

The second study was recommended to evaluate the dosage form proportionality within a particular product

line. Three treatments were chosen to represent the low, middle and high ends of the product line and each treatment was administered as a single 600 microgram dose under fasting conditions. Pharmacokinetic analyses again, as with the other study, were conducted using total thyroxine without an endogenous baseline correction.

Finally, the issue of formulation which is, in my opinion, perhaps the most important aspect of this guidance. It's a small section in the guidance, but it has a very big impact. In order to be acceptable to the agency, a sponsor's products must target 100 percent of the label claim, something that had never been done before. Unaccountable or stability overages were viewed as unacceptable and would prevent the approval of that product.

Between June 1999 and July 2001, nine sponsors submitted stand-alone NDA applications. The first product was approved in August of 2000. There are currently six approved levothyroxine sodium tablet NDAs, and I have them listed here. We've got Lloyd, Jerome Stevens, Genpharm, Jones, MOVA, and Abbott Pharmaceuticals.

I'd like to conclude by saying that the process that I've just described has had a major impact in improving the quality and consistency of these six FDA-approved products. Important issues, such as overages,

content uniformity, and bioavailability, have been addressed, and product-specific dissolution tests -- I'll repeat that again because it's very important -- product-specific dissolution tests have been conducted. And it's very important that these were specific to the product because it allows for lot-to-lot consistency and quality evaluation.

These steps go a long way in addressing some of the historical concerns that were brought up earlier with levothyroxine sodium tablets.

Thank you.

I'd like to introduce Drs. Wartofsky and Granneman from Abbott Laboratories.

DR. WARTOFSKY: I'm Leonard Wartofsky. I'm Chair of Medicine right here in Washington at the Washington Hospital Center, Professor of Medicine at Georgetown University. I'm here as a consultant for Abbott, and I have also received honoraria from virtually every other levothyroxine manufacturer for speaking.

For 25 years, I was at Walter Reed Army Medical Center and am now at the Hospital Center, and I've been in leadership positions in the ATA, the American Thyroid Association, and the Endocrine Society. But I'm a practitioner of endocrinology, seeing thyroid patients every day.

I'd like to stress that the FDA recommendations you've just had reviewed to determine bioequivalence are not sufficiently sensitive to detect the small differences in thyroxine levels and their physiologic effect that we clinicians are concerned about. These small differences have a significant clinical impact on both safety and efficacy.

T4, as you've heard, is the synthetic version of the naturally occurring thyroid hormone. There is no substitute for thyroxine. All our patients require lifelong therapy and the medical community relies on thyroxine as being truly bioequivalent.

The decision of the committee here today is extremely important because 13 million Americans rely on thyroxine.

You've heard a little bit about TSH this morning. I'd like to review it some more. Here is the pituitary gland that makes and releases TSH, appropriately in the center of the slide. It stimulates the thyroid gland to release T4 and T3 which circulate in the blood, binding to tissue receptor sites where the metabolic action of thyroid hormone is exerted. There's negative feedback back to the pituitary and the hypothalamus turning off TSH. So because we cannot look at all of these other tissue levels effectively, TSH is our window into the body where

we can judge the effectiveness of a given level of T4 or a given dose of levothyroxine and its physiologic effects.

So we physicians use the TSH level to individualize our patient doses of thyroxine and optimize those doses and clearly, as you heard this morning, small changes in a dose can cause significant clinical effects. Like Dr. Tuttle, who you heard this morning, I specialize in thyroid cancer and it's very important for my patients to have their TSH levels exactly titrated to where we want it. The manufacturers facilitate this need of the clinician by providing 12 different dosage strengths. Differences as little as 9 or 10 percent between these doses can make a big difference for our patients.

You heard also this morning of entities of mild thyroid failure or mild hyperthyroidism. In these entities, the serum T4 levels, either free or total, are normal or within the reference range, but in the case of mild thyroid failure, the TSH is slightly elevated, in mild hyperthyroidism, the TSH is suppressed. These two entities are a model and correlate exactly with our patients who are taking exogenous replacement thyroxine.

The importance of these slight differences are illustrated by this study that you've seen already twice this morning. This was a study by Carr in the U.K. that looked at a group of hypothyroid individuals and optimized

the perfect thyroxine dose judged by their serum TSH levels and TRH tests, as well as thyroid hormone levels and a symptom questionnaire. They then increased the dose or decreased the dose by 25-microgram increments or decrements and you can see the major effect on TSH with a slight reduction or suppression with a slight increase, and these are over the range of again the various dosage strengths of levothyroxine that are available to us.

This has an impact on particular populations in our practices. Most patients taking thyroid hormone tend to be older because of the increased frequency of hypothyroidism with each advancing decade, and our older patients have cardiovascular disease, particularly sensitive to excess thyroid hormone. You heard from Dr. Brown this morning about the risk of hypothyroidism on the neonate, on the newborn, and pregnant women who are under-dosed with thyroid hormone will give birth to children with lower IQ, and you've heard about the importance in our patients with thyroid cancer. With insufficient dose of even a mild degree, cholesterol levels go up, atherosclerosis is accelerated, leading to an increased risk of heart attacks, myocardial infarction, as well as the risk in the newborn I've already mentioned.

My concern is that the current assessment of bioequivalence is not adequately sensitive to detect these

small differences that matter. These are the real concerns and experts need to decide on a new approach that will address these concerns. Anything less, such as continuing the current bioequivalence standard, would be a disservice to we practicing physicians and our patients.

I'd like to turn it over now to Dr. Granneman who will demonstrate how the current bioequivalence criteria perpetuate presumptions of bioequivalence that create the potential for the adverse clinical consequences that you heard about from all of the physician speakers this morning.

Thank you.

DR. GRANNEMAN: I'd like to thank the FDA and the committee for inviting us to talk about the results of our study and various baseline correction procedures.

Although we will spend a lot of time talking about the ways that you can correct for endogenous T4 products, there's a larger question that we have to consider. Ultimately, we have to ask the question, does bioequivalence translate into therapeutic equivalence? When we look at the new guidance that the FDA has proposed, we fear that with the current criteria, this may not always be the case and that, as a result, there will be some patients who are at risk.

I'm going to give you an abstract of the study

that we ran and then go through the details of the study, but basically, if you don't correct for endogenous levels of T4, then you cannot detect differences of 33 percent in dose. All the correction factors work actually quite well in terms of detecting 25 percent differences in dose, but they're unable to detect 12.5 percent differences.

Beyond that, we looked at some other factors and found TSH particularly good and promising for distinguishing very small differences in dose in bioequivalence studies.

Shown here are the results of our study. This is Study 417. It was a typical randomized, three-way crossover comparing doses of 600, 450 and 400 micrograms. The difference between 400 and 600 is 33 percent. All these doses came out of the same lot of Synthroid.

Going to the bottom, the FDA has proposed a certain scheduling sampling routine and what we did in our analyses is to go well beyond what they have proposed. Instead of just looking at three samples prior to dosing, we characterized the entirety of day minus 1 and then rather looking out to day 2, we took our sampling all the way out to day 4. Rather than looking at just T4, we looked at T3 and TSH because we have been told that TSH is very critical in assessing the action of thyroid hormones.

Now I'll tell you a little bit about the

correction procedures that we used in our study. First, these three curves here are for those three very different doses and just looking at the curves, you can see they're very, very close to each other, very little difference between the three curves.

Now, to go to the various correction procedures that one might envision using, first, there's the horizontal correction. The premise behind horizontal correction is that that large exogenous dose of T4 has absolutely no effect on endogenous T4. In other words, there's no perturbation of the biology by that large a dose.

The next correction procedure takes just the opposite approach. It says that that large dose totally and completely shuts down the production of endogenous T4. So what's left in the body washes out with a half-life of 7 days.

What are the other approaches? One, we know that biology isn't that constant like the horizontal correction method, that there's fluctuation through the day. So what we did was use day minus 1 data and corrected based on that.

Then we had a rather novel approach. Since we collected TSH in the study and since we found that TSH was suppressed, why not marry the good parts of the last two

correction procedures and make the wash-out dependent on the suppression of TSH? That's this method? I showed two different curves. Actually this allows every individual to be corrected. So if there's very little suppression of TSH, then it comes very close to the day minus 1 method.

Then the last thing we did, as recommended in the open session, TSH is a factor that has to be looked at and we did in our study.

Now, in this graphic to the right, I'm going to show the results of our study. Just to orient you, down at the bottom of the graph, what we're going to plot is the area under the curve ratio for a 450 microgram dose versus 400. The regulatory goal posts of 80 to 125 are shown in the yellow lines. The magenta vertical line is unity. Now, since we're comparing 450 versus 400 microgram doses, that appears right here, the blue line and it goes vertically. So what we want to do is to look at how well the point estimate and the confidence interval center about this blue line because that's reality.

We're going to ask four questions of the methods that we looked at. The first question is, will the method detect 25 percent differences, a rather large difference?

And then in the open session, many of the physicians said that, really, it's critical to be able to

detect 10 or 12.5 percent difference in dose. So what we have are three questions associated with that. First, is 1.125 within the confidence interval? Does it hit this blue line? Second, is the difference between those two doses statistically significant? And third, will the test fail that difference?

To go to the results, if you don't do any correction, then everything fails and there's really very little more to be said about that.

Now, this is what we understand to be the FDA preferred method of horizontal correction. What we find is that that procedure can detect 25 percent differences but cannot detect 12.5 percent differences.

The next method we looked at, 7-day half-life, it's about the same. There's a little bit of improvement in the point estimate but still not very good.

The day minus 1 correction method actually does a little bit better. The point estimate is migrating toward the real value and the confidence interval now contains the true value.

And last, the TSH method that takes into account TSH suppression does even better, and a new thing appears in the statistic in that the difference now becomes statistically significant between those two doses. But those two doses would still be declared to be

bioequivalent.

Now, at this point, let me just focus on a couple of things that have already been mentioned before with a couple of the other speakers.

First, looking at the confidence intervals, they're quite narrow, and as was mentioned before, this is a narrow therapeutic margin drug. So with such low variability and narrow confidence intervals, do we really need these regulatory goal posts of 80 to 125 when we're thinking about consumer risk?

Next, the TSH correction method. It gets four checkmarks. It finds the two doses to be different from each other, but it has some disagreeable characteristics. Number one, it's more sensitive. Actually the point estimate is above the true value, and also the other issue that was talked about by Dale is the confidence interval is relatively broad. So if you were to use TSH alone, then you would have to seriously consider broadening the confidence interval.

Now, back to the issue of horizontal correction, a picture was drawn with a perfectly flat line with horizontal correction. Well, in reality, these are data from day minus 1 in our study for the three periods and the curves are not perfectly flat, and in fact, at 18 hours, there's a significant decline in levels. So when

you use the perfectly flat horizontal correction method, you're making an error due to that data point.

Another thing that we noticed in this study that is a testimony to the complexity of the biology of T4 kinetics is that with successive periods, this is period 1 in green, period 2 in magenta, period 3, the baseline is dropping, despite the fact that it's more than 7 weeks since that last dose. So we've affected the kinetics of endogenous T4 by giving those very large doses.

Well, the biology of T4 is very, very complex and this is a schematic that sort of is a testimony to that complexity. I'm not going to go through that schematic, but I want to make a point that Dale mentioned.

In the discussion of bioequivalence, there's talk about rate and extent of absorption and appearance of the active principle in the biophase. Well, what we're talking about here as the biophase is the tissue compartment and the active component probably is more T3 than it is T4. It's much more active in binding the thyroid receptor. Well, of course, we can't measure T3 within cells, but we have a very good surrogate of that, and as has been spoken to before, that surrogate is TSH.

A thing that I have to make a point about is that all of these pathways in this diagram, all of those arrows are controlled by the levels of T3 and TSH. As a

result, the half-life of T4 can be as small as 4 days in hyperthyroidism, as much as 9 days in hypothyroidism. So it changes. It's a moving target.

And the other thing that was mentioned, TSH changes exponentially with small changes in T4, and in fact, you can have a doubling in TSH for only a 12.5 percent change in T4.

Now, consider the biostudies. Consider normal volunteers. T4 is a very, very unusual drug. Unlike other drugs, if there's too much of it on board, then its clearance increases. If there's not enough of it, then its clearance decreases. Think about that in context of a biostudy when you're administering two non-equivalent doses. The body is going to try very hard to get rid of both of them, but it's going to try harder to get rid of the larger dose.

In the briefing document, I've shown you a graph of what happens to TSH. I'm going to show you a little bit of a different orientation about TSH response. We're going to express T4 and TSH as a fold change from baseline in our biostudy. We're going to invert the TSH ratio because TSH and T4 are reciprocally related.

So these are the results for those three doses. The thing that you can notice with the high dose, 600 micrograms, the ratio is 1.7, in other words, a 70 percent

increase, and then looking at the two lower doses, they're superimposable.

Now, the question is, how does TSH respond to these relatively small perturbations in T4? That's shown here. It's a very dramatic change. The point we want to make here is for a very small perturbation in T4, TSH is excellent in distinguishing small changes. There is pronounced hysteresis, but the bottom line is that TSH is a very good discriminator and it adds biologic context. After all, why are physicians using TSH in their management of patients?

Going back to the horizontal correction procedure, this is a typical dose that I've simulated here. The red line is what we expect is happening to endogenous levels based on a NONMEM fit. Here's the horizontal correction procedure there in blue.

The points that we can make here is that it's biologically inconsistent. The baseline is probably not flat and it's not variable.

If you use this procedure, you've reduced the true area by 10 to 15 percent and that will result in attenuation of differences between non-equivalent formulations.

There are two other characteristics that we really need to think about with this correction procedure.

One is it produces negative area under the curve values, and second, the imputed half-life is only 2 to 3 days, whereas we know the real half-life of T4 is about 7 days. So there are some issues with the method.

To summarize our study, all the correction methods are good for 25 percent differences. They're not good for 12.5. The horizontal correction method does have some biologic inconsistency. We know the intrasubject variability in T4 is low. We know it's a narrow therapeutic margin drug. If we are to be serious about detecting 12.5 percent differences, then the standard 80 to 125 criteria are probably too broad for T4. In using TSH, you get more discrimination.

Now, there are many physicians who don't understand or don't trust bioequivalence. What they really want to know is if you can switch two products and pose no risk to the patient.

Another option to think about in biostudies is if we have a problem with correcting for baseline, why not get rid of the baseline? Why not study the drug? Why not study bioequivalence in subjects that don't have any thyroid function? There's precedence for this for estrogen products. The study would have to be a multiple dosing. It would have to be steady state, and you would really like to validate it with known differences.

Now, what marker to use? Well, physicians use free T4. They also use TSH. If we were to use those, though, you would have to define the maximally accepted changes in TSH are to ensure the physicians of their therapeutic equivalence.

So to conclude, small differences matter. Products that differ by 12.5 percent cannot be detected with the current criteria, and we fully believe that we should bring all the scientific prowess in academia, FDA, endocrine societies, and industry to consider the issues of how to construct proper evaluation of bioequivalence in these T4 products.

That concludes my presentation.

DR. JOHNSON: Well, this part of the presentation will now focus on the FDA's current recommendation for evaluating levothyroxine sodium bioequivalence. However, before I begin, I want to make a couple of comments with regard to some of the slides that we just saw from Abbott Laboratories.

First of all, we want to thank Abbott Laboratories for conducting their correction method study. This data was confirmatory and very useful when the FDA decided to adopt a baseline correction method for evaluating levothyroxine sodium tablet bioequivalence.

However, there are some drawbacks with this

particular study design. The use of 400 and 450 microgram doses yielded thyroxine concentrations that were closer to baseline. This is problematic because it prevents an accurate evaluation of the true differences that exist between the two doses and this is likely due to some sort of baseline interference. That's why the agency has recommended in the guidance and continues to recommend that doses of 600 micrograms or greater are used.

Also the checkbox slide that compared the different evaluation methods clearly shows why TSH on its own is inappropriate. The point estimate was detecting a 24 percent difference when in actuality there was only a 12.5 percent real difference between the products.

Now on to the bioequivalence design. This is the current study protocol that we're recommending to sponsors seeking A-B ratings. A single-dose, two-way crossover study in which healthy subjects will receive 600 micrograms of both test and reference product. Pharmacokinetic analysis will be conducted using total thyroxine with a baseline correction.

Now, let me discuss some of the rationale behind the study design. First of all, the use of healthy subjects allows us to do a single-dose study and a single-dose crossover study is the most sensitive method for evaluating the true formulation differences between

products and that's really what we're looking at. A single-dose study cannot be conducted in patients. A 600 microgram dose in healthy subjects provides concentrations that are significantly higher than the individual subject's baseline T4 values, and the farther away from the baseline that you actually get, the more accurate the evaluation of the products. The issue of nonlinearity is really not an issue since the subject is receiving the same amount of drug in each treatment period.

Regarding the bioequivalence measures that have been discussed this morning, total thyroxine is the preferred measure for demonstrating bioequivalence. It can be accurately measured in vivo and is the drug that is being administered to the subject. T3, on the other hand, is merely an active metabolite, and the Food and Drug Administration does not use active metabolites for conferring bioequivalence, unless the active parent cannot be measured in vivo.

Finally TSH. TSH is a biomarker and it's an indirect measure. It's downstream from what is being administered and it's considerably more variable than thyroxine. It's also very easily influenced by other environmental factors, such as time of day and ambient temperature.

To kind of give you an idea of where each of

these measures fits into this negative feedback system, let's start with the lower left-hand corner, with the L-T4 or T4 inputs. Once you have conversion to T3, the T3 has an inhibitory effect on the hypothalamus which ultimately results in a reduction in the amount of TSH secretion from the anterior pituitary, but this is not a mutually exclusive event. As mentioned before, other factors influence the TSH values.

According to the Code of Federal Regulations, in descending order of accuracy, sensitivity and reproducibility for determining bioavailability and bioequivalence of a drug product, the best choice for evaluating bioequivalence is the concentration of the active ingredient and that's where T4 fits in. TSH, on the other hand, would be relegated to the third or fourth category.

As was made very clear in the previous presentation, using total thyroxine without a baseline correction is insensitive for conducting bioequivalence studies with levothyroxine sodium tablets and the FDA completely concurs. Rather, a baseline correction method whereby the mean of three pre-dose samples is subtracted from all of the subsequent post-dose samples. This is the preferred method and it is adequately sensitive for evaluating levothyroxine bioequivalence.

Now, when the agency decided to adopt a baseline correction method for bioequivalence, we went back to data from the six original NDA applications. Dosage from proportionality studies from four the six NDAs were re-evaluated using the baseline correction method and they're presented here.

Let me orient you to this slide. On the left-hand side, we have four products, 1, 2, 3 and 4. The first two columns are AUC and the second two columns are Cmax. This is a three-way crossover study. The dose that was used for the comparison was 600 micrograms, and as you can see, the bioequivalence criteria, when they're applied to these data sets, the confidence intervals still fall well within the confidence bounds of 80 to 125.

These results also show the power and sensitivity of this method because it shows the sensitivity to detect real differences as evidenced by the values circled in red. We've got a 14 percent increase in level 4, in product 4, for AUC, and on the same scale, we also have about a 9.5 percent decrease. The confidence limits, if this were slightly more variable, would have clearly failed.

In conclusion, the FDA has thoroughly reviewed each of the NDA applications that have come in. We've had a lot of data -- there were nine submissions -- the

literature and the recent correction methods study, and we've concluded the following. Levothyroxine can be evaluated in healthy subjects. A single dose crossover study is a preferred method for detecting the true differences between products. T4 is an appropriate and sensitive measure for this particular process, and a baseline correction method using the mean of three pre-dose samples is adequate when determining bioequivalence between two levothyroxine sodium products.

Thank you.

I'd now like to introduce Dr. Barbara Davit who will be speaking on potassium chloride.

DR. DAVIT: Thank you. I'm Barbara Davit, and I recently became the Deputy Director for the Division of Bioequivalence in the Office of Generic Drugs.

I'll be presenting some information today about baseline correction methods for endogenous compounds for which the Division of Bioequivalence has a fair amount of experience and that's potassium chloride.

I'll be discussing the design of potassium chloride bioequivalence studies that we've been implementing, the application of baseline correction methods to bioequivalence study data, the impact of baseline correction on bioequivalence study outcome, and to accomplish this, I have two cases to present, one in which

baseline correction made a difference in study outcome, the other in which it made no difference in study outcome. Finally, I'll compare two methods for baseline correction to determine if the method of baseline correction made an impact on the outcome of the bioequivalence studies.

We recently revised and updated the guidance for industry on bioequivalence testing of potassium chloride products, and the web address is given here. The guidance describes recommendations for study design and emphasizes special dietary considerations to achieve a stable potassium baseline. The guidance also discusses collection of urine samples to evaluate pharmacokinetics and finally methods for data analysis.

To help in establishing a stable baseline that contributes minimally to the amount of potassium that we measure after giving a dose, we recommend that study subjects eat a diet with a controlled potassium intake. Normal potassium intake ranges from to 50 to 100 milliequivalents a day. Thus in these studies, the recommended potassium intake is on the low end of what's considered a normal diet for potassium intake. It's not really a low potassium diet or a diet deficient in potassium but rather a controlled potassium diet.

Fluids are given according to schedule. Bioequivalence of potassium chloride products is determined

by giving subjects a single 80 milliequivalent dose and, finally, to determine the baseline, we take urine samples during two days before the dose is given.

This schematic summarizes the study design for the potassium chloride bioequivalence studies. The basic design is a two-period, two-sequence, two-treatment crossover with each study period 8 days in duration. The controlled potassium diet is given throughout the study. The diet is given for 4 days. Then on study days 5 to 6, urine is collected at various intervals throughout the day. Dosing takes place on the morning of day 7 and then urine is collected again at various intervals throughout days 7 and 8. The urine collection intervals on days 5 and 6, the baseline days, match the urine collection intervals on days 7 and 8, the post-dosing days.

I mentioned that we collect urine to measure potassium excretion in these bioequivalence studies. As has been discussed earlier today, most of the time in our bioequivalence studies, we measure drug concentrations in plasma, serum, or blood because this is the most sensitive and accurate way to determine bioequivalence. However, in the case of the endogenous substance potassium, urine measurements give the most accurate assessment of bioequivalence.

Now, this is in part because when potassium is

absorbed, most of the absorbed dose is excreted through the urine, but also it's because, as Dr. Conner brought out earlier, serum potassium is a very insensitive measure. This is because body homeostatic mechanisms maintain serum potassium concentrations within a very narrow range. The normal range for serum potassium concentrations varies from 3.5 to 5 milliequivalents per liter.

We noted that in typical bioequivalence studies of potassium chloride oral dosage forms, serum concentrations increase by only about 5 percent after a single dose of 80 milligrams. What this means, recalling the schematic that Dr. Conner showed earlier, is that the baseline in serum is a very high amount relative to the increase that's observed following a dose. Therefore, measuring potassium in serum will not give an accurate measurement of bioequivalence of two formulations because the additional potassium in serum after dosing is a very small amount of the total.

In evaluating bioequivalence of potassium chloride oral dosage forms, we asked that firms calculate these parameters: the amount of potassium excreted in each collection interval, the cumulative excretion over 24 and 48 hours, the maximal rate of excretion, and the time of maximal excretion. We asked that firms report both the baseline and the uncorrected data, but the bioequivalence

statistics are performed only on corrected data.

The key parameters for bioequivalence evaluation are the cumulative amount of potassium excreted in the 24-hour interval after dosing and Rmax, which is the maximal rate of excretion. The 90 percent confidence intervals for the ratios of test to reference must fall within the 80 to 125 percent goal posts.

We asked that baseline correction be subject- and period-specific. So in other words, what this means is that the amount excreted in the 24-hour interval after dosing in urine is corrected by subtracting the average amount excreted in 24 hours and determined during the two pre-dosing days.

Rmax, the maximal rate of excretion, is corrected by subtracting the baseline from the corresponding interval averaged from the two pre-dosing days, and as an example of this, how we would ask firms to do this, consider subjects from whom Rmax occurred from 6 to 8 hours after dosing. So if Rmax was observed during the interval corresponding to 1 o'clock to 3 o'clock p.m., then the correction would be done by subtracting the rate of potassium excretion from the baseline days that was observed from 1:00 to 3:00 p.m., and as I said earlier, it's subject- and period-specific.

Baseline corrections are done for potassium

chloride drug products because we'd like to determine, as accurately as possible, the amount provided in the dosage form. The baseline reflects the amount of potassium provided in food. So we assume then, after dosing with potassium chloride tablets, the amount of potassium in urine excreted above and beyond the daily amount due to food is due solely from that which is provided from the drug product. Thus, the amount of potassium provided from the two formulations can best be determined by doing the baseline correction which would correct for the amount of potassium excreted from food intake.

This figure shows the 24-hour excretion rate in a typical bioequivalence study of potassium chloride tablets. The figure is a plot of the excretion rate versus the midpoint of the urine collection interval, and the plots are from test subjects in period 1, reference subjects in period 1, test subjects in period 2 and reference subjects in period 2. There's a small amount of fluctuation during the day and this may be due to meals or it may be due to circadian rhythms or a combination of those. However, as you can see in the figure, the 24-hour baseline is consistent from period 1 to period 2 and in the test and reference subjects.

So the first case study that I'll discuss I'll call formulation A, and it's for a 20 milliequivalent

extended release tablet product. For this particular product, without baseline correction, both the amount excreted over 24 hours and Rmax met the 90 percent confidence interval criteria. However, with baseline correction, Rmax, the maximal rate of excretion, did not meet the 90 percent confidence interval criteria. Therefore, we found the application unacceptable.

This chart shows the 90 percent confidence intervals and point estimates for the amount of potassium excreted in the 24-hour interval after dosing for formulation A. The ratios for this parameter fell within the 80 to 125 goal post for the 90 percent confidence intervals. However, with baseline correction, the 90 percent confidence interval was wider than with uncorrected data.

As I mentioned earlier for this particular product, formulation A, without baseline correction, the test-to-reference ratios for Rmax, the maximal rate of excretion, fell within 80 to 125. When we did the baseline correction, the lower bound of the 90 percent confidence interval for Rmax was outside of the 80 to 125 range.

Then what we did was we compared two different methods of baseline correction to see if there was a difference in the results. We subtracted the mean excretion rate from the corresponding interval and that's

the usual way of correcting for potassium chloride excretion, as I discussed earlier. We also subtracted the overall mean excretion rate from the 2 baseline days and we found that the outcome was the same, regardless which of the two baseline correction methods we used.

This figure shows the potassium excretion rate plotted versus the midpoint of the collection interval time. The upper plots are for uncorrected excretion rates after dosing for both the test and the reference. The lower plots are the excretion rates pre-dosing. The baseline excretion rate contributes about 20 to 30 percent of the total excretion rate.

This figure shows the potassium-excreted rate corrected for baseline, plotted against the midpoint of the post-dosing collection intervals and it's for the test product versus the reference product. This is for formulation A, the product that did not pass bioequivalence criteria for Rmax, and you can see here that the differences in Rmax are more apparent after correcting for baseline than before correcting for baseline.

The second example that I'm going to present is also for a 20 milliequivalent extended release tablet product. For this product, both the amount excreted in 24 hours in Rmax passed the 90 percent confidence interval criteria whether baseline correction was done or not and

this particular generic product, therefore, was found to be bioequivalent to the reference product which in this case was the K-Dur microburst tablet.

For formulation B, the amount of potassium excreted in urine in 24 hours after dosing passed the 90 percent confidence interval criteria with or without the baseline correction. However, as we've seen earlier, the 90 percent confidence interval was wider after baseline correction than for uncorrected data.

We also compared for formulation B two different ways of baseline correction for Rmax. As previously, we compared the effect of subtracting the mean baseline from the 2 baseline days versus subtracting the mean baseline from the corresponding collection interval, and the test-to-reference ratios for Rmax were within the 90 percent confidence interval criteria whether corrected or uncorrected and regardless of which correction method was used. However, as I've mentioned previously, the confidence intervals were wider when baseline correction was used.

So finally, to conclude, we have found that baseline correction is essential for evaluating bioequivalence of potassium chloride tablets, and we've also found that the correction method as proposed in the guidance for industry is reproducible during the two study

periods. We found that baseline-corrected data are more sensitive to differences in formulation performance than uncorrected data. We've also found that baseline correction can make a difference in whether a product passes or does not pass the 90 percent confidence interval criteria, and finally, we found that although it was essential to do a baseline correction of the two methods that we tested, the method did not affect the study outcome.

Thank you very much, and now Dr. Conner will summarize this afternoon's presentation on bioavailability and bioequivalence of endogenous substances.

DR. CONNER: Again, to restate some of the technical problems or questions or, I guess you could say, controversial issues with endogenous substances in general, some of the things that we've discussed or seen illustrated are assay sensitivity. If you have a very small amount of something, especially after baseline correction, it's important to be able to give your assay the best chance at measuring the signal and to be able to get the best sensitivity from that. So one of the ways you do that is to give a dose that's large enough to give a good signal, if you're measuring in plasma or any other bioassay.

Endogenous baseline, as I mentioned before, feedback inhibition is always something that you need to

deal with as an issue. Different variations or circadian rhythms, what you saw illustrated, and whether it has linear or nonlinear pharmacokinetics.

So again, I feel like I harp on this endlessly, but again, the core question in bioequivalence is one of formulation. So you have to always keep that in mind, that you're really looking at how that manufacturer has made their formulation and how the results of that work actually perform when it gets into the in vivo situation. Sometimes we lose track of that core question with other very legitimate clinical concerns about how this is used and how the drug or drug product actually works.

But the BE question is a very simple and should be a very directed one on what is the best way of looking at those two formulations, whether it be the same manufacturer making changes in their formulation, whether it's questions between whether two lots are indeed far enough away to cause clinical problems or whether it's looking at a generic product or a substitutable product from another manufacturer.

The question is always back to how have they made that formulation, how successful have they been in controlling both the variability in the performance of the formulation, as well as whether that formulation hits its target or the performance characteristics that that

manufacturer, the formulation designer is going for. So we generally look at the performance in basic as the release of the drug substance from the drug product.

As I said before, I think we can all agree the drug substance has to get out of the drug product to be able to get into the body and create a therapeutic effect, and based on regulations of what we're instructed to do and on good science, we're looking at both the extent of release or the extent of availability from any formulation as well as how quickly it happens or the rate.

We saw a couple of examples where baseline correction -- or there is an endogenous baseline, one of the characteristics of endogenous substances. And the question is how to best account for that baseline? Does it need to be subtracted from the data that you're measuring? If so, how do you go about doing a proper subtraction or proper baseline correction? You have to really look at a variety of different things, characteristics of the baseline, various methods for correction, you saw some illustrated in previous talks, and what I think is very important is magnitude of baseline in relationship to the total values that you're measuring.

If you really think it through, something with a very, very small baseline in relationship to the total amount after a dose has very little effect on your eventual

outcome, and you can go through some calculations to prove this to yourself.

If you look at something, on the other hand, like potassium chloride, where that baseline is a very large percentage of what you're seeing as your signal when you measure it in plasma or blood, actually subtracting that baseline would probably mean that virtually no study that you did, even on a product against itself, would probably be likely to pass. I mean, it becomes so sensitive and the signal becomes so small, when you subtract most of that signal away, that certainly two lots of the same product would be unlikely to pass if you did that study with any kind of reasonable number of subjects.

So on the other end, any tests you do should both discern the differences that you're interested in, yet not fail products that are almost if not identical. I mean, that's an unreasonable test if you fail a product against itself.

So the magnitude of the baseline is a characteristic when you look at a new drug substance or a new endogenous substance, that you really have to look at. Is it worth subtracting a baseline if it's extremely small and has little effect on the results or, on the other side, if the baseline is extremely large, is there any way that I can subtract that baseline out and still get any kind of a

reasonable test? So those are the two extremes.

Obviously, it increases the difficulty of accounting for the baseline if there are feedback mechanisms, as there are with most hormones, that change the baseline with differences in doses or differences in blood levels. So that becomes a significant problem in how best to construct a baseline subtraction scheme when you have a feedback mechanism.

So finally, I guess it's not really a question but kind of an end point is that when we look at new endogenous substances, can we develop a thought process or a decision tree, if you will, of various factors that are important in determining how we're going to deal with that particular substance? Do we or do we not subtract baseline? How are we going to measure it? At what dose? Is it going to be even possible to use our normal, I think, well-accepted and reliable plasma concentrations or are we going to have to go to yet another scheme or another area of measurement to try and develop an understanding about bioequivalence methods that are going to assure that those products behave in an equivalent manner?

So that's the endpoint that we're looking for as an overall scientific construction of thought about how to approach these products, how to look at the various variables and characteristics of a new endogenous product

and how to construct a proper way to do formulation comparisons.

DR. KIBBE: I guess now is a good opportunity for those of you who have been taking copious notes on the presentations in sequence to ask questions. Wolfgang is smiling at me. Marv will start.

DR. MEYER: First of all, I'd like to compliment Abbott as FDA has done. So oftentimes, we have the innovator company whine about differences, perceived differences, imagined differences, extrapolated differences, simulated differences, and they never come in with real data. So I think Abbott has done a good job of trying to gather some data, and I personally appreciate that.

A couple of questions I have. It seems to me, in my non-endocrinology background, that TSH is much like measuring blood pressure. A clinician might like to see changes in blood pressure and an endocrinologist might like to see changes in TSH, but if you can show what's going on with a drug you're administering, given an appropriate baseline correction, it seems to me that that's the appropriate thing to do.

I'm a little troubled by repeated reference to 12.5 milligrams as being critical to patient therapy, and I didn't see any data. Now, there may be a lot of physicians

know that that's true, but the data in the literature all seems to revolve around the Carr study. And Dr. Wartofsky showed the Carr study and had arrows inserted for a 12.5 percent change but really didn't show any data. It was just kind of if this would happen, then this would happen.

If you look at the Carr study, the original document in 1988, the only relevant comparisons, I think, in terms of changes in TSH with changes in levothyroxine dose are the ones that go from 150 to 175, which is a 17 percent change, and 175 to 200, which is a 14 percent change. Everything else is 20 percent or greater.

And in that context, I'm trying to move toward the 12.5 percent change and there's no data for that, but there's at least a 17 percent change. There's only 3 patients out of the supposedly 21 that were in that category that had changes from 150 to 175 or 175 to 200. The 1 patient that went from 175 to 200, which is a 14 percent change, didn't seem to have much of a change in TSH. The other three seemed to have some changes. So that's basically 3 subjects out of 21.

So I wonder how serious the issue is that the Abbott study was not able to detect a 12.5 percent difference. If that were a 12.5 percent difference in other drugs, we'd say, well, the system worked. So that's an open question. I'll leave that to perhaps somebody more

knowledgeable on thyroid therapy than I.

Plus, the Carr study, there are always questions about compliance. They did tablet counts, but whether that worked or not, there was no -- since that was an '88 study, we don't really know. They obviously went from one strength to the other in order to get the different strengths. There's no information available on content uniformity or potency as they moved to the different strengths.

I guess one substantive comment might be out of the Abbott study, the comment on a carryover, and I didn't hear much discussion of that. I know in the old days, FDA would fail a study if it had a carryover, and then they kind of backed off of that and said, well, if you can justify it or there's no reason for the carryover, it'll be okay. Is that still an issue, and should we be concerned about apparent carryover in the levothyroxine?

DR. KIBBE: That's a lot of questions. Is anybody jumping in here with answers? Go ahead.

DR. LESKO: Thank you, Art.

We've seen the Carr study about three or four times today, and I think there's some points in that article that need to come on the table for consideration.

First of all, TSH is not a blood pressure. Blood pressure is a surrogate endpoint for clinical

effectiveness and blood pressure has been correlated with mortality and morbidity. TSH has not been correlated in any prospective study that I'm aware of with clinical symptomatology of thyroid disease.

If you look at the Carr paper very carefully, it's probably the lowest evidence of clinical studies that we would consider; that is to say, it's not a randomized, double-blind study. It's not even a randomized study. It's a case-control study and certainly that has merit, but it also has many limitations and weaknesses.

It's also an artificial study in that optimal doses were obtained after thyrotropin-releasing hormone injection. In other words, it was a simulated TSH response to an exogenous injection of TRH.

But as I read through that, there were a couple of points that the authors made that I thought were interesting. An optimal dose was determined for each patient. However, in 2 patients, more than one such optimal dose was evident, so these were not unique optimal doses. In 4 patients, no dose tested resulted in a normal TRH response, and the optimal dose was taken to be that dose at which the TRH response was closest to normal. So that's at least 30 percent of the patients in whom a normal dose was not successfully achieved.

I think importantly, though, no significant

differences were observed in any clinical symptomatology, weight, pulse rate or any clinical index over the range of thyroxine doses that were studied, 25 micrograms below or 75 micrograms above the optimal. No patients receiving doses from 25 micrograms below to 75 micrograms above optimal were considered to be hypothyroid or hyperthyroid.

As you get to the discussion part, the authors comment that these data highlight the relative insensitivity of clinical observations which fail to detect clinical differences between patients receiving thyroxine at various doses within the range studied. In other words, there's no connection between the TSH and the clinical observation.

Patients actually felt better when the thyroxine dose was increased to 50 micrograms above the dose required to normalize TRH response. The authors attribute that to a placebo effect, but there's no evidence that that's the case.

Finally, at the end, the authors conclude that our study does not address the all-important question of whether the TRH test fulfills the criteria of a gold standard, whether its application would yield optimal clinical results with minimum morbidity. The value of routinely adjusting thyroxine doses according to any test of thyroid function remains controversial.

Well, it still is controversial because I did a more recent search of the literature, and I think we need to consider the current status of thyroid function tests, and there was a series of articles in the British Medical Journal that looked at this. They talked about the confusion surrounding thyroid function tests, and they cited two studies of recent vintage, studies in 1,580 in-patients, 630 out-patients, found that thyroid function tests performed as a screening test yielded abnormal results in 33 and 20 percent of patients, respectively. In both studies, these biochemical tests suggested thyroid disease incorrectly. They gave false positive results in 9 out of 10 cases.

So the TSH, as I understand it, is a biochemical test designed to help in the diagnosis of a thyroid disorder. I'm not so sure it's an adequate test for the demonstration of bioequivalence, and I think one of the presenters talked about a range of TSH that would be adequate for bioequivalence. Well, I guess I would take a step back and say based on the literature evidence that we have for the TSH as a measure of dosing and its relationship to clinical outcome is certainly controversial. I would imagine that the confidence interval on that would have to be really quite wide, but I'm not sure how you would establish it. There are no

clinical studies.

This is from the British Medical Journal, July 2001. The TSH test, currently the most widely-used blood test to diagnosis thyroid dysfunction, is an unreliable test of thyroid function that has no proven scientific biochemical basis. Anecdotal evidence indicates that the biochemical diagnosis of hypothyroidism with the TSH test is very poorly correlated with the clinical diagnosis of hypothyroid symptoms. Free T3 and free T4 are reliable evidence, etc.

So I guess the point of bringing this all up is that while we've talked about TSH as unequivocally a measure of therapeutic outcome, I think it still needs to be looked at very carefully because certainly the literature is conflicting with what we've heard today, and I think we need to look at it more closely.

DR. KIBBE: Thank you, Larry.

Wolfgang?

DR. SADEE: Yes. I have some concerns about TSH measures to assess bioequivalence, and although I do not doubt that it's probably one of the better measures to titrate a patient, what we have to consider first is the relationship between the dose and the effect. And in this case, it is a very steep dose-response curve and that was already alluded to by their saturation phenomena, but also

the steepness of the curve implies that very small changes cause very large changes in the TSH level and the coefficient, which is a measure of how steep the curve is, is probably up to 5 or 10 as an exponential.

What that means is that the measure of TSH is extraordinarily sensitive, as was pointed out by many of the speakers earlier, but sensitivity does not mean accuracy. It does not convey an idea as to really what the bioequivalence is. It may be the ultimate desire to achieve this, a certain level of TSH, but it cannot measure the dose necessarily, and what we have to ask ourselves -- and this is really the question I'm coming to -- is, what are the main variances or differences?

To me, the greatest difference is in different patients that will provide the biggest difference. The next one may be different formulations, then different batches of the same formulation, and different times, the changes over time within the same patient. That may be in the same order of magnitude in terms of a variance to the others.

So if we design our tests that are extraordinarily sensitive to small changes in the dose and that's granted, I do think that's truly the case, it may fail many of the formulations, whereas the more important aspect is what is the variability within the same

formulation, etc.

So I think the TSH test is useful clinically, but it may not be the proper test for establishing bioequivalence. Do you have some comments to that?

DR. KIBBE: Anybody?

DR. CONNER: I pretty much agree with you. I'll defer to Steve's specifics about levothyroxine, but I think anything with a steep dose-response curve -- if you looked at the depiction of the confidence interval on TSH, number one, the point was made that the point estimate was way off of what it should be. So number one, you weren't even getting the right answer from the center part or the mean.

But also if you look at the breadth of that confidence interval which is a reflection of variability, I would tend to guess that if you did that study on two lots of any manufacturer's product, it would probably fail, if that study was done, with that level of variability.

In fact, I would even go out on a limb and say that you might fail testing if you took the same lot and just randomly divided it into two sections and studied it in a crossover fashion and did the same study, you would have a pretty decent chance of failing identical stuff from the same lot, given that study and that level of variability.

So even all other things aside, if you just looked at that level of variability of your response, you would either have to study lots of subjects or you would have to increase the confidence interval limits a substantial amount to have a reasonable test.

DR. SADEE: So would you agree then that if we to apply TSH tests to compare different formulations, then it should also be done for complying different batches of the same formulation?

DR. CONNER: I won't agree to that.

DR. KIBBE: I think one of our guest presenters might have a couple of comments, and we'll give him a chance to --

DR. WARTOFSKY: Really speaking for myself as a clinician and not for Abbott, I have to take exception to some of the comments that were made.

What you heard this morning were hundreds of years of clinical experience from senior members of The Endocrine Society and the American Thyroid Association, seeing tens of thousands of patients and seeing the importance of these minor 12.5 microgram differences that were alluded to.

The Carr study has been criticized. It's not an optimal study, I would agree, but it is one of the only ones we have. The importance there was that TRH was not

used to stimulate TSH. TRH was just another test assessing the physiologic level of those patients. They were looking at TRH tests. That was not really the criterion.

There is indeed a well-established correlation of the extent of clinical disease, hypothyroidism, with TSH elevations. It's as evident as that high blood pressure causes strokes and heart attacks. It hasn't been studied because it's so self-evident to endocrinologists.

And the differences that were alluded to in some of the studies, yes, TSH will vary and thyroid function will vary, and it depends on whether we're talking about acute administration or chronic. It's a matter of dose and duration. A 12.5 microgram difference in thyroxine over years will cause atrial fibrillation, subclinical hyperthyroidism, and osteoporosis. It may not create a big problem over the course of a 6-week bioequivalence study, but long term for our patients, it does. We know there are data on how many times we physicians have to change the dose by 12.5 micrograms to make our patients feel better and be less symptomatic. There are data that can be provided for that.

So we're talking about a TSH test that may not be perfect but it's the best thing we have now, and what we're asking the committee to do, what I'm asking the committee to do is to consider getting the experts

together, analyze all these pros and cons and come up with what would be the best method of assessing bioequivalence because we don't have it.

In reference to Dr. Johnson's comments, the choice in the Abbott study to me of 600 versus 400 versus 450, that wasn't the design of the study. That study, as far as I can tell, was designed to assess whether we could detect differences between 10 and 30 percent, not whether we should assess bioequivalence using 400 or 450. That was not the intent.

It may not be that TSH may not be best, but certainly T4 is not good. He alluded to changes that can affect TSH. All the same things can affect T4. T4 is affected by upright posture. It's affected by fluid changes. It's affected by protein binding. Many more things than TSH is. TSH can be measured both sensitively and accurately. The variation in a good TSH assay is extremely tight. We have third- and fourth-generation TSH assays that make that irrefutable.

Dr. Johnson, I think, ignored the wealth of the data this morning, the Hennessey data, that showed that T4 levels could be the same but TSH is not. The pituitary is not sensing those levels as the same, and even if, in his last slide where the confidence intervals in the bioequivalence test between the four preparations did fall

between the 80 to 125 standards, that's not being questioned. It's whether that standard really reflects bioequivalence in the pharmacodynamic sense. To us physicians, it does not. It may be good pharmacokinetics, but it's not pharmacodynamics and that's what we're concerned about, not the statistics but the clinical effect.

Thank you for the opportunity to make some comments.

DR. KIBBE: Gary, do you have anything?

DR. HOLLENBECK: Well, I'm not sure now is the best time to ask it, but I am somewhat intrigued by the question that was asked about doing these studies in patients with no thyroid function.

Could someone from FDA just sort of respond and answer that question? Is that an unrealistic thing to do?

DR. JOHNSON: Yes. Actually, we've talked about that quite a bit within the Clinical Division and we felt that that was an unrealistic study type, just to do it in athyrotic patients. We need to do, first of all, the recruitment process, and second of all, if we're taking into consideration TSH, the number of subjects would be astronomical. So the decision was made actually prior to 1997 when this first guidance was put together.

DR. HOLLENBECK: I wasn't referring to TSH. I

was just referring to testing a traditional bioequivalence test using patients with no thyroid function. So is the first part of your answer the really relevant one here, that there aren't enough subjects to do that?

DR. JOHNSON: We did not feel that there were enough subjects to do that.

DR. KIBBE: Do we have anybody else who has any questions?

(No response.)

DR. KIBBE: No other questions?

DR. MEYER: While Dr. Johnson is there, the recommendation on one of your slides was baseline correction based on three pre-dose rather than across the whole profile, and you said data provided by Abbott. Is that correction 1?

DR. JOHNSON: Yes, it is.

DR. MEYER: Although the correction 1 seems to give better point estimates, less close point estimates in correction 3.

DR. JOHNSON: Which --

DR. MEYER: Correction 3 is where they correct for the whole profile.

DR. JOHNSON: Right. The 24-hour.

DR. MEYER: Right.

DR. JOHNSON: There is some variation within

the day on the baseline. There's some diurnal variation. It tends to be under 10 percent per individual in the individual, and when you compare taking intensive sampling over 24 hours and compared that against the mean of three pre-dose samples, it's not very much different. I think it's 7.77 versus 7.75 percent CV. So we didn't feel that it would be necessary to do that.

The other thing in that study, it was a point-by-point subtraction method, and the fact of the matter is we still don't know exactly what happens to baseline on treatment, and it doesn't make sense to increase your noise because the point estimates switch and the confidence intervals change.

DR. MEYER: I guess I was just looking at the AUC 96 hours. For a 1.125 difference in dose, the point estimate is 1.08 for the correction method 3 and 1.03. So there was a 5 percent improvement, if you will, by using the overall correction.

DR. JOHNSON: Right, and we attribute some of that improvement to the fact that when we're comparing the 400 and 450 microgram doses, you are getting closer to baseline and that noise from the baseline is going to interfere with that evaluation. That was the point that I was trying to make.

DR. KIBBE: Ajaz has a few comments.

DR. HUSSAIN: No. I think just in closing, this was sort of a general discussion on endogenous drugs, and I think Dale provided sort of a framework for moving forward with decision tree criteria.

The question I think I have in my mind is, as we move forward to this, does the committee feel that a decision tree criteria would be a valuable step in terms of dealing with these compounds because we will have a number of endogenous substances to deal with? The list that Dale provided, this partial list, I think the numbers are quite high, and I think we'll have to deal with every one on a case-by-case basis. But is there a framework of a decision tree that could evolve from this discussion?

DR. KIBBE: Pat?

DR. DeLUCA: Yes. I have a question just to go back on that, and I noticed when Dale was talking, he seemed to be talking about bioavailability and bioequivalence, and are we mixing things here? It seems like with the endogenous substances, bioequivalence may be something difficult to determine. The patient is the critical factor here, and what we have here is certainly something that's pharmaceutically equivalent and bioavailable, but beyond the bloodstream, can we really assess the bioequivalence? It just seems like it's going to be a horrendous task to try to do that. That's a

clinical marker.

DR. SADEE: It would appear to me that endogenous substrates are so different from each other, that making the decision tree in which you force how you proceed might be very difficult. I think it would have to come up with a decision tree and then we can test it against all the endogenous substrates that we might want to look at. The example of thyroxine is one. It's such an extreme example, although the elements are all there, the self-regulation and so on, but you may take it on a case-by-case basis, but if you do produce a good decision tree that people can be actually guided by, then it would be very helpful. We need to see the details.

DR. HUSSAIN: So from that sort of comment, should I perceive that we may not take this up as the first topic in the Biopharmaceutics Committee and move to something else then?

DR. MEYER: I haven't had a lot of time to think about prioritizing which of the 12 topics you gave us.

I agree with Wolfgang. I mean, can a decision tree be developed? I haven't the foggiest at this point, but I think it's a worthwhile exercise to crystalize your thinking, and if it turns out you can't, then you can't, but if you can, it's helpful.

DR. KIBBE: Part of, I think, the assignment of topic priority order is also how close is the flame to the -- I mean, if this is something that the agency needs to move on and move on quickly because there's a lot of patients at risk, there's a lot of issues at hand, then even though we might like more development time before we really get into it, I think we need to start looking at it in that light. If there's a lot of window of opportunity to be leisurely and take our time, then maybe not.

I agree with Wolfgang. I think coming up with a decision tree that works for every compound isn't going to work. Coming up with a model of a decision tree that might apply different concepts might work, and when I start to look at the model and start to get it in my mind, I might be even happier with it.

DR. HUSSAIN: The decision tree was intended to sort of take us to different approaches to address different issues and how to make those decisions, where to go sort of thing.

DR. VENITZ: I would be very much in favor of you pursuing looking at a decision tree. Just food for thought. In my mind at least, there are mechanistic things to consider that relate to our understanding of the underlying biology as we heard today, and then there are more empirical things. How do we baseline correct? Do we

need to baseline correct? What's the contribution of endogenous versus exogenous? So I do think it's perfectly worthwhile to do so.

DR. KIBBE: Does anybody have anything else? We're scheduled for a break at 3:00 to last till 3:15 and it is 3:17, which means that you sacrificed your break. No. I'll give you all 10 minutes, and we'll get back and we'll ask Ajaz to make up for the time when he does his presentation.

I would like to meet with Barbara Davit for a couple of seconds.

(Recess.)

DR. KIBBE: Ladies and gentlemen, fellow scientists, colleagues, clinicians, media reporters, and others, we need to get started again, and we are fortunate in that we have speaking to us near the end of the day Ajaz Hussain without slides.

DR. HUSSAIN: I think what I would like to do is first again thank all the speakers, especially the physician community, which came to this meeting to share their concerns and perspectives with us. I think from my perspective, they are our customers and I think we have to give very careful attention to their concerns, and we will continue to do that. I think customer satisfaction is paramount, and I think without customer satisfaction, you

can't build confidence and generate trust. So that is, I think, a key challenge that we have, and I will use that as a framework for the following section of this discussion.

I had mentioned earlier, Helen asked me to take the lead for the Therapeutic Inequivalence Action Coordinating Committee. What that is, it is a committee that looks at consumer complaints. It looks at reports of inequivalence that come to the agency through many different means, through publications, scientific literature, and all sort of sources. Clearly, the discussion we had fits into that, and I think we always have to carefully review every aspect of every complaint and come to some resolution.

But at the same time, I think dealing with perceptions also is a challenge, and it's a very difficult task to separate perception issues from actual science and technical issues and that's clearly a big challenge for us.

For that purpose, I think, and for other purposes, what we have done is we have created a Rapid Response Team, which was actually created in the year 2000, to deal with burning issues that need to be addressed quickly through lab-based or other scientific support functions. We use this Rapid Response Team to actually get to a root cause as quickly as possible, using scientific data. Nakissa will talk to you about that team and share

with you some examples.

So that is a part of the research program that I have kept at the OPS level. We have an Office of Testing and Research, but the Rapid Response Team sort of brings all the resources available to us and all of our offices to deal with issues in a very rapid manner. So you'll hear Nakissa talk about that.

But there are other research programs at the office level, and at some point, I'll make this committee aware of those programs in much more detail, and I think it's an exciting program that we have on computational toxicology. FDA has probably the best database available on drugs in terms of their safety, efficacy, and a number of things, and if you don't utilize this database effectively, then you're not doing the right thing and you're not learning from the database that we have.

So there's a group within our office which has developed excellent predictive models of toxicology using data that is available to us in our submissions, and many of these software products are available commercially now through a collaborative research and development agreement. So these are structured activity-based, bioinformatics-based predictive tools that we have been developing, and we will be expanding some of the scope of this to drug-drug interaction and other areas too. But that will be for a

different advisory committee that will bring this information to you.

With that, I'll ask Nakissa to come and share with you what's the Rapid Response Team and what is it doing.

DR. SADRIEH: Hi. I'm going to talk about the Rapid Response Team. This is the last presentation of this advisory committee meeting, and I promise it's going to be a short one. Thank you.

I'll give you an overview of the Rapid Response Team. It's a research-based mechanism that helps provide research support to the review divisions and ultimately the drug approval process. It also helps to respond to literature reports of drug inefficacy or toxicity. The Rapid Response Team is also used to evaluate suspected causes of therapeutic inequivalence. By this, I don't mean that we go and sort of do detective work to find out what's the cause, but when a cause is identified and research needs to be done, then the Rapid Response Team sort of mobilizes the laboratory resources to try and address the research needs to come up with an answer for what's been identified. Also, we've provided some data for counter-terrorism initiatives, and I'll be talking about those.

As Ajaz mentioned, the Rapid Response Team was created in November of 2000, and I'd like to point out that

the Rapid Response Team and rapid response project is only a small part of the research that's done under OPS. There's a lot of research that OPS does. This is a very specialized aspect of it, where all the various resources are basically mobilized to take care of specific projects. So I don't want you to think that this is everything that OPS does for research.

The function of the Rapid Response Team is to provide timely and specific research support, whether it's laboratory-based or literature-based, for designated regulatory issues that require further agency study, and the goal is basically to provide review divisions with sound scientific data which may be used in the regulatory process.

What is the Rapid Response Team composed of? It's basically a group of multidisciplinary scientists from all the offices under the Office of Pharmaceutical Sciences; namely, the Office of Testing and Research, the Office of New Drug Chemistry, Office of Clinical Pharmacology and Biopharmaceutics, and Office of Generic Drugs. Initially, the Rapid Response Team was under the Office of Testing and Research, but now it's been placed in the immediate office of the Office of Pharmaceutical Science and the purpose for that was to increase the breadth of the types of research studies that are done and

to bring also some more visibility to the types of projects that we actually do do.

Some of the projects -- I'll go into that later, but the Rapid Response Projects are in general very high-priority projects and they have a short turnaround time. We decided that a maximum of six months is what we'd like to set for the completion of the studies, and they're expected to have regulatory impact, direct regulatory impact. By that, I mean they should support reviewer recommendations, whether in the Office of New Drugs or the Office of Generic Drugs. They should support labeling changes, and they should support advisory committee issues.

Some of the examples of some of the past projects that we've done, we've done palatability studies of doxycycline and potassium iodide -- these were two separate studies -- in human subjects to identify dosing regimens that would be appropriate for pediatric populations in the event of a bioterrorism incident.

Another study which was looking at the permeability of commercially available gloves to lotion and shampoo that was used in the treatment of lice.

We do routinely studies for dissolution properties of select drugs. I cannot mention the specifics about the drugs because some of the data is proprietary and it's about applications that are still pending.

We do determination of BCS classification of select drugs, and another study that we have done is looking at the neurotoxicity of ketamine in juvenile animal models and this was an interesting study. Ketamine is used in children to set bones when they break their bones, and there were reports in the literature that ketamine may be neurotoxic. That was in an animal species, in the rat, and our labs were able to duplicate the data and show that in fact it is toxic in rats. So the study has now been expanded, and the National Toxicology Program has actually taken that up and they're going to be looking in a non-human primate model to try and see if the neurotoxicity is actually present in that model or not. This could have significant regulatory implications.

The resources that we have at our disposition are all the laboratories within OTR and those include the Laboratory of Clinical Pharmacology, the Laboratory of Pharmaceutical Analysis, which is located in St. Louis, the Division of Product Quality Research, and the Division of Applied Pharmacology and Research, and in addition to that, we also have contracts set up with several universities, including the University of Tennessee and the Uniformed Services University. The work that we did with the palatability studies, for example, was done by the University of Tennessee.

In fact, right now, we're working on another palatability study and that's the palatability of ciprofloxacin tablets in human subjects, again to identify appropriate dosing regimens for pediatric populations in the event of a bioterrorism incident. Again, this is because the national stockpile has only got solid oral dosage forms, and it's important just to know if we can actually prepare a solution from these solid oral dosage forms that would be palatable for children to take in the event of a bioterrorism incident.

Other on-going projects are to support the Therapeutic Inequivalence Action Coordinating Committee which Ajaz mentioned, the TIACC, and they've kept us quite busy, too.

We are also working with the Office on Drug Safety on some data mining projects to characterize adverse event profiles for generic drugs as compared to innovators. So this is a literature-based research study.

We're also providing laboratory support for select RSR projects, and RSR projects are review science research projects that are specifically sponsored by reviewers and so we support, not all of them but some of them, in trying to get their studies done.

What have we accomplished? Well, we've generated some data for publication on the FDA website

called The Home Preparation Procedure for Emergency Administration of Potassium Iodide Tablets to Infants and Small Children. I have the website there, if you're interested. We've also generated data to update drug labels.

Where do we plan on going in the future? The hope is to provide sound scientific data which may contribute to policy decisions by regulators, and we also would like to identify new areas of regulatory research which might help policy development. We would also like to collaborate with scientists outside of FDA to identify new technologies which might be incorporated in the drug development process.

Thank you very much. I said it was short.

DR. KIBBE: Thank you. Wow.

There's got to be at least one question. Efraim, you're back. You can ask a question.

DR. SADEE: I have a quick question. Are the adverse effect or the side effect studies available on line? Do you make this information available or the data mining --

DR. SADRIEH: On the data mining, yes. We just started that. It depends on what we get and we have to look at that, but if it's data that's out in the public domain, it will definitely be published and it will be

available to everybody. But it's an exciting project and we hope to get some good results from that one.

DR. HUSSAIN: Just to add to that, I think when we get reports of therapeutic inequivalence, for example, or side effects, generic versus innovator, our databases right now are not truly optimum to find the signal and to see whether the signal is real or not, and the study that Nakissa is planning to do is to go back and look at select drugs where the endpoint for either the adverse event and so forth are well defined and see whether we can start taking signals of differences between generator and innovator, and based on that maybe, hopefully, construct a better database to be very proactive in looking at these signals, hopefully in real time, later on.

DR. KIBBE: Anybody else?

(No response.)

DR. KIBBE: Thank you.

DR. SADRIEH: Thank you.

DR. KIBBE: Ajaz, are you going to end?

DR. HUSSAIN: I'll be very short, and I think everybody's tired, and again I think the two days, plus many of you have attended the third day of the training session, we really appreciate your time and effort, and as you sort of get to understand the advisory committee -- and I hope this meeting was really helpful to expose you to the

different types of challenges we face on a daily basis in FDA and the struggle and how to bring science into it -- I think your advice and your input becomes very valuable for us to keep moving forward in the right direction and hopefully keep improving the science of what our regulatory policies are based on.

I think the two observations that I would like to make over the last two days, and the two observations I had for today's discussion, I think one was the manufacturing issues in terms of when we say that quality cannot be tested into a product, it has to be by design. I think that is an area that we need to discuss a bit more because, for example, one of the aspects that we discussed was what happens if there is one unit has no drug or one has more drug, and how does the current system avoid that. I think that is the concept of quality by design or quality being built in. You cannot design a test to find that, unless you test 100 percent of the lot. So the process validation, the science of process validation is essentially what allows us to move in that direction and so forth. So I think that is something we will have to discuss at length and as we move forward with other methodologies.

Again, I think the endogenous substances and the challenges you see in terms of customer satisfaction

and the customer's physician, the challenge ahead is tremendous. You can imagine in the sense of how do you build confidence in a generic drug program when customer satisfaction is a challenge. And I think I will really need your help as we move in that direction, how to do that. Clearly, we have a lot of work ahead of us in trying to sort things out and clearly define the issues and explain the processes that we adopt and the science that we have to our customers, not only the patients but the physicians and the pharmacists out there.

I think on day one, I think the key issue that is in my mind is the topical products, whether they are topical products for skin. I think many of the issues are also customer issues and customer perceptions on quality of generic drugs. So we struggle with pharmaceutical equivalence there and now we struggle with bioequivalence. So how do you define therapeutic equivalence? I think the key there which also is quite apparent is when you're trying to evaluate differences in formulations that were designed to be similar, where the differences are actually minimized by design, then what sort of test do you use to say the difference is not big enough when the test may be far more variable than the differences that you see in the products you're testing? That's the struggle that is inherent in this discussion and that was apparent on both

days, and so how do we articulate our position not only to the physicians and pharmacists but also the customers will be a big challenge for us.

With that, I think again I thank all of you for your patience and your advice and we'll take this seriously and at the same time, all the comments we have received from the public, we'll take that into consideration and work towards the next advisory committee.

Thank you.

DR. KIBBE: I'd just like to thank Ajaz, Helen, and the rest of the FDA staff for doing the best they can to make us comfortable and productive and being here with the right answers and all the help.

I also would like to thank all my colleagues who contributed and spent a lot of their time here to help the agency and, through the agency, the health and welfare of the American public. You should go home proud of yourself for having made that sacrifice and not frustrated on having not accomplished as much as you want.

I look forward to seeing you all again.

(Whereupon, at 3:53 p.m., the committee was adjourned.)