U.S. FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

+ + + + +

EQUIVALENCE OF LEVOTHYROXINE SODIUM PRODUCTS
JOINT PUBLIC MEETING
(Cosponsored with the American Thyroid Association, The Endocrine Society, and the American Association of Clinical Endocrinologists)

+ + + + +

MONDAY, MAY 23, 2005

+ + + + +

The joint meeting was held at 8:30 a.m. in the Boardroom of the National Transportation Safety Board, L'Enfant Plaza, Washington, D.C., Dr. David G. Orloff of CDER and Dr. Paul W. Ladenson of Johns Hopkins University moderating.

FDA REPRESENTATIVES:

DAVID G. ORLOFF, M.D., Director, Division of Metabolic and Endocrine Drug Products
DALE P. CONNER, Pharm.D., Division of Bioequivalence
BARBARA M. DAVIT, Ph.D., Division of Bioequivalence
ERIC P. DUFFY, Ph.D., Division of New Drug Chemistry
STEVEN K. GALSON, M.D., M.P.H., Acting Director, Center for Drug Evaluation and Research
ROBERT LIONBERGER, Ph.D., Office of Generic Drugs
HENRY J. MALINOWSKI, Ph.D., Office of Clinical Pharmacology and Biopharmaceutics

ALSO PRESENT:

PAUL W. LADENSON, M.D., Johns Hopkins University School of Medicine
JAMES V. HENNESSEY, M.D., Brown Medical School
E. CHESTER RIDGWAY, M.D., University of Colorado School of Medicine
STEVEN I. SHERMAN, M.D., University of Texas M.D. Anderson Cancer Center
LEONARD WARTFOSKY, M.D., M.P.H., Uniformed Services University of the Health Sciences/Washington Hospital Center

PUBLIC SPEAKERS:

BETH BRANNAN, Sandoz
GREGORY BRENT, M.D., Secretary, American Thyroid Association
ROSALIND S. BROWN, M.D., Lawson Wilkins Pediatric Endocrine Society
ALFRED ELVIN, Ph.D., Sandoz
ALAN P. FARWELL, M.D., American Thyroid Association
LISA H. FISH, M.D. The Endocrine Society
JEFFREY R. GARBER, M.D., Secretary, AACE
IRWIN L. KLEIN, M.D., New York University School of Medicine
ROBERT A. JERUSSI, M.D., Jerussi Consulting
MICHAEL J. LAMSON, Ph.D., King Pharmaceuticals
HOWARD LANDO, M.D., practicing endocrinologist
WILLIAM H. LANDSCHULZ, M.D., Ph.D., Abbott Pharmaceuticals
JOHN LEONARD, M.D., Abbott Pharaceuticals
PETER LURIE, M.D., M.P.H., Public Citizens' Health Research Group
ERIC POMERANTZ, Sandoz
ROBERT RICHARDS, M.D., Louisiana State University Medical Center
SALLY SCHIMELPFENIG, Sandoz
FRANK SISTO, Mylan Pharmaceuticals
BRUCE WEINTRAUB, M.D., Trophogen, Inc., formerly National Institutes of Health
LAWRENCE C. WOOD, M.D., Thyroid Foundation of America
CHERRY WUNDERLICH, Thyroid Cancer Survivors' Association

A-G-E-N-D-A

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

8:44 a.m.

MS. CUNNINGHAM: Okay, let's try again. There are just a couple of administrative announcements I would like to make. There are three sign-in sheets for the public comment periods that start after the first break. Well, after lunch, 12:50, 2:15, and 4:05. There's no food or drink allowed in the auditorium, but if you want to bring something, take a snack or something, there is a room back there that you can sit in. There is a screen there also. Would you please turn off your cell phones and your Blackberries as it interferes with the uplink and causes static on the lines. The restrooms are located in the lobby, and we have a really ambitious schedule, and we're already behind schedule. So would you please keep to your allotted time. I have a timer here that I will set. It will stay green, it will go to a 2-minute warning where it turns yellow, and then when your time is up it turns red, and the floor opens up and takes you.

Now, I'd like to turn the podium over to Dr. Galson. He's the Acting Director for the Center for Drug Evaluation and Research at the Food and Drug Administration. Dr. Galson?

DR. GALSON: Thank you, Rose. Thank you for all the hard work that you and your colleagues have done putting together this meeting. I wanted to welcome all of you to our Public Meeting on the Therapeutic Equivalence of Levothyroxine Sodium Drug Products. The meeting today is cosponsored by the American Thyroid Association, the Endocrine Society, and the American Association of Clinical Endocrinologists. We appreciate very much the opportunity to further explain FDA standards and methodology for determining levothyroxine sodium therapeutic equivalence.

These products came on the market, as you all know, over a half century ago without FDA review and approval for safety and efficacy. Although the efficacy of levothyroxine products was demonstrated in scientific literature, over may years, we received reports of wide deviations in stability and potency that raised FDA's concerns about the quality of the products used in clinical practice. As a result of this concern, in 1997 FDA declared that oral levothyroxine sodium drug products were considered new drugs and would be required to obtain marketing approval under new drug applications. Applicants would be required to demonstrate that they could

consistently manufacture a high-quality product of predictable potency and stability over the shelf life of the product.

Since that announcement, FDA has approved seven new drug applications for levothyroxine products. Although none of these was originally rated as substitutable for another product, which is what we call AB rating, we have now approved supplemental new drug applications and generic drug applications from sponsors who demonstrated the therapeutic equivalence or interchangeability of their products with certain others.

As we made these regulatory decisions, some, including members of the societies that are cosponsoring this meeting today, have questioned our methodology for assessing bioequivalence, which is a confirmatory test in FDA's determination of interchangeability of drug products, including levothyroxine products. Some have expressed concerns that patients are being harmed by involuntary substitutions of levothyroxine sodium products. Let me assure you that patient safety is FDA's number one priority, and we believe that the decisions that we've made with regard to levothyroxine sodium products are in the best interests of the patients and of public

health. Our purpose in agreeing to cosponsor this meeting is to help you to better understand our rationale and methodology so that members of the thyroid community will be able to prescribe any of the approved products with great confidence and assurance of patient safety.

I'm sure you've all read about our latest safety initiatives in FDA, which include making our regulatory decision-making processes more transparent. Our willingness to cosponsor this meeting is furtherance of that patient safety goal. This meeting will include formal presentations by FDA and by representatives of the cosponsoring societies. We also intend to provide as much time as possible for comments by other interested parties during the open discussion sections of the agenda. Again, let me thank all of you for the opportunity to be here today, and to contribute to this important discussion.

At this point I'd like to turn the podium over to Paul Ladenson who's the president of the American Thyroid Association and a professor at Johns Hopkins, as well as the coordinator for the societies at this meeting. Welcome, Dr. Ladenson, thank you.

Dr. LADENSON: Well, thank you very much Steve, and thanks in general to FDA for its

willingness to move ahead with this workshop. I want to first of all thank the National Transportation Safety Board where we are reassured that anything that moves runs more smoothly than things that are static.

I want to thank first Dr. Janet Woodcock whose vision more than two years ago was that we hold this workshop at which we could have a thoughtful and thorough and I hope open-minded and transparent discussion of the methodologies currently in use and the concerns that many hold about them. I also want to thank Dr. Galson, whose integrity and tenacity have ensured that this meeting did go forward after long delay. And finally, to thank Dr. David Orloff whose collegial cooperation has been essential in putting together the format and content of today's meeting. So from the societies' perspective, the American Thyroid Association, the Endocrine Society, and the American Association of Clinical Endocrinologists, we hope that today's discussion will be thoughtful and thorough, and that it will be only a beginning in continuing the process of improving the precision of thyroxine therapy. So thank you Steve and David.

I also am the first speaker, and so I will just shift gears, having already been introduced, and the topic of my presentation, which I will think will

permit us to catch up some of the time we've lost, is simply to introduce you to levothyroxine sodium as a widely employed and narrow therapeutic range drug. Our society's concerns at the outset, and openly, are that current bioequivalence standards, when combined with current prescribing and dispensing practices in the United States are inadequate to ensure the safety of thyroxine-treated patients. We think that working together we can all do better, and we think we must do better, especially for certain vulnerable populations to which you'll hear reference during the course of the day, patients who rely upon great precision in thyroxine therapy, pregnant women and their growing children, the elderly, other individuals with vulnerabilities of their heart and skeleton to modest degrees of thyroid hormone excess and deficiency, and especially thyroid cancer patients whose titration with thyroxine therapy need be especially precise.

And our goals, the societies' goals in today's meetings are to instigate a commitment to four measures that we think can take everyone to the next step in precise thyroxine dosing: more stringent standards for bioequivalent testing, the use of TSH as a pharmacodynamic measure, stricter regulation and label warnings regarding the switching between

formulations, and the requirement for re-titration which you'll hear later today as being widely ignored, and finally to amass data to instruct each of these preceding steps to undertake a properly designed definitive crossover clinical trial to assess the real therapeutic equivalence of thyroxine formulations, a trial that would include appropriate controls and measurement of a TSH as a pharmacodynamic index.

There are some unique challenges of thyroxine as a drug that everyone in this room is intimately familiar with. This is a compound which using TSH principally as a surrogate is known to have adverse effects at both ends of its spectrum. And you'll be hearing from later speakers about some of these effects. We don't intend to belabor them because Dr. Orloff and I agreed early on in our planning for this session that we would stipulate all agree that levothyroxine therapy entails a very narrow therapeutic index of efficacy and safety. Indeed, the FDA has spoken to this point, saying that levothyroxine sodium is a compound with a narrow therapeutic range where small differences exist between therapeutic and toxic doses. And further define generally narrow therapeutic index drugs as substances that are subject to therapeutic drug

concentration monitoring, and/or where product labeling indicates a narrow therapeutic range designation.

In fact, the FDA has been even more specific in its communication with levothyroxine manufacturers about what our societies agree is one appropriate precision point. In 2001, FDA said that a 9 percent refill to refill difference could have serious consequences for thyroid patients. More recently, FDA approved thyroxine products with dose increments as little as less than 9 percent, for example, the 137 microgram versus 125 microgram thyroxine tablets. And just last year, FDA said that its standards will not allow products that differ by 9 percent or more in potency or bioavailability to be rated therapeutically equivalent.

Levothyroxine is also a challenge because it is an endogenous substance with a plasma protein-bound pool of hormone. Residual thyroid gland function is the rule among patients who are treated with thyroid hormone for hypothyroidism and sometimes that function is autonomous, complicating therapy. This residual endogenous function can interfere with bioequivalence test data in normal subjects, and FDA has recognized the importance of the large endogenous

thyroxine pool, and its endogenous production by altering its approach to bioequivalence testing with baseline correction, although that's not been fully codified in its communications with manufacturers.

We believe, the societies, that there is evidence that current bioequivalence standards are inadequate, and that that evidence arises from two broad sources. First, clinical experimentation, and you will hear later this morning from Dr. Hennessey about clinical trials in which different doses of a known single formulation of thyroxine have escaped detection or exclusion using current bioequivalence standards. We are even more concerned, however, about the reality of a regulatory performance over the past year and a half. This shows you data just posted approximately a week ago on the FDA's site examining the actual application data of test products compared to reference products. You'll see that one of the most widely employed novel products, when substituted for one of the most widely prescribed thyroxine brands is associated with a difference that is significantly above 9 percent. Indeed, among the approved products, you can see that in every case one of the 95 percent confidence limits exceeds the 9 percent narrow therapeutic index goal that FDA itself has set

forward.

Now, we're blessed in a sense by the precision of the hypothalamic-pituitary-thyroid axis, which in itself instructs us about the importance of precise thyroxine dosage in physiology, and enables us by measurement of TSH concentration therapeutically to adjust thyroxine therapy. We know from a study that you will hear quoted, I am sure, a number of times later today, the Carr Study, that modest changes in thyroxine dosage among patients who have been, as in this study, carefully titrated to optimal TSH concentrations can result in either over-treatment or under-treatment. Within this study, 25 microgram increments resulting in 88 percent and 55 percent of patients having TSH concentrations that fall out of range, and have been associated with adverse clinical consequences.

Now, with TSH measurement, it should nonetheless be a piece of cake for clinicians and patients to adjust thyroxine appropriately. Clinical experience, though, in this country and overseas suggests that this really is not a reality. You see here four studies, one from Parle, British General Practitioners, Canaris, a population-based study performed in Denver, Hallowell data from the NHANES

III series, and Ross from the august thyroid clinic at the Massachusetts General Hospital showing a remarkably consistent phenomenon, that from 15 to 20 percent of thyroxin-treated patients, even in specialty practices, and certainly among broader populations, are over-treated, 15 to 20 percent under-treated based upon TSH as a surrogate marker associated with known adverse clinical effects.

When one thinks about the complexity of thyroxine therapy, it is perhaps no surprise that this kind of variation occurs. From the delivery of raw drug with known purity and strength to manufacturers, the production of drug, its distribution and storage, all of these steps are carefully monitored by FDA. Then we have the role of the physician in prescribing drug accurately, the patient's filling of the prescription, the pharmacist's dispensation of the drug appropriately responding to physician's direction, the patient's role in storing the drug and using it for an appropriate period of time, and then perhaps most importantly in this sequence of events adhering to therapy and taking the drug as prescribed. Drug absorption, and in the case of thyroxine therapy its activation by deiodination in target tissues also are subject to physiological and pathophysiological

changes. And drug interactions, just as they interfere with absorption, can also alter the metabolism and clearance of thyroxine, a phenomenon that can also be affected by physiological changes such as pregnancy and aging.

As we think about any such complex sequence of events, how does the variance of each individual phenomenon relate to the whole? And this is a simple equation that describes that relationship. Here, perfection in terms of dose-prescription versus dose-received. A variation in a single parameter, such as bioequivalence, or adherence to therapy, interference with absorption or metabolism resulting, as you can see, for an individual patient taking a typical dosage of thyroxine of perhaps a 10 to 15 microgram per deciliter per day difference. There is no guarantee that the variance in a single step, for example, the shelf life of a medication, will cancel out other variances. And as you can see here, when you add imprecision in other steps, this potential variability becomes even greater, with the possibility of a perfect storm of variance alterations that could result in serious clinical consequences for a patient.

Every day across the country physicians caring for the 13 million Americans who take

levothyroxine make the kinds of dose adjustments that you see illustrated here on this slide, often changes, indeed in the majority of cases, changes that are less than 25 percent, and often less than 12.5 percent in their magnitude. The concern of our societies is that these changes be made with deliberation and precision, and not be made -- or not be countermanded by chance.

So in conclusion, and introduction to today's meeting, FDA and clinical sub-specialists have improved the precision of thyroxine therapy for the Americans who need it. Nonetheless, we believe that current pharmacokinetic standards, when combined with the reality of contemporary prescribing and dispensing practices, are not adequate to ensure the safety of patients taking thyroxine, or the efficacy of thyroxine therapy in some cases. We think we can do better, and we think we're obliged to work together to do better, especially for the vulnerable populations that I mentioned at the outset of my talk.

You're going to be hearing from four speakers during the remainder of the day representing our societies. Dr. Hennessey, who will talk further about our concern and recommendations regarding the stringency of bioequivalence standards. Dr. Ridgway, who will talk about TSH as a pharmacodynamic measure

to augment our assessment of levothyroxine products and their therapeutic equivalence. Dr. Wartofsky, who I think will provide you a window on the reality of contemporary practice, and the need for stricter regulation and label warnings regarding the switching between formulations, and the inadherence to the re-titration requirement that is so widespread. And then finally Dr. Sherman is going to dream with you a bit about what a properly designed, definitive crossover trial would look like to assess the equivalence of thyroxine formulations, including use of TSH as a pharmacodynamic measure. So again, I want to thank Dr. Galson, and thank Dr. Orloff, and like the rest of you, I look forward to our thoughtful and thorough discussion of this issue through the remainder of the day. Thank you.

DR. ORLOFF: Thank you, Dr. Ladenson. Our next speaker is Dr. Dale Conner. He's the supervisory pharmacologist from the Office of Generic Drugs in the Center for Drug Evaluation and Research at FDA. You can't hear me? We'll work on it. Dr. Conner.

DR. CONNER: Can you hear me? Okay. Today I'm looking forward, as I'm sure most of you are, to a very stimulating discussion, a very lively one. However, it's my job that I've been assigned to

give the introductory material to explain the basics of this pharmacokinetically-based bioequivalence technique that we use on literally hundreds, if not thousands, of products in both the NDA or new drug arena, as well as in the generic drugs arena.

So first off, you can look all through the literature and other places and find a variety of different definitions of bioequivalence, some fairly loose and broad saying that virtually any formulation of any type can be compared to another. When I talk about bioequivalence for the purposes that we're discussing today, I'm talking about pharmaceutical equivalence whose rate and extent of absorption are not statistically different when administered to patients or subjects at the same molar dose under experimental conditions. So I'm using a very tight and very specific definition of bioequivalence.

And the first important point of this is when we look at substitutable or switchable products that are eventually granted an AB rating, we're always looking at pharmaceutical equivalence. And what we mean by pharmaceutical equivalence is a tablet is equivalent to a tablet. In our system, a capsule is not equivalent to a tablet. So that would not be given a switchable or AB rating.

Pharmaceutical equivalence also has the same amount of the exact same drug. If we talk about two different salts of the same drug, we're not talking about pharmaceutical equivalence. So it has the same dosage form, intended for the same use, and it has the same amount of the exact same drug in it. So a suppository is not pharmaceutically equivalent to a tablet, and so forth. So that's very important for our definition and what we're talking about now. And I think probably everyone understands that all of the products at issue here are all tablets containing the same nominal dosage strengths of levothyroxine.

Why do we do this? First and foremost, the purpose of conducting bioequivalence studies is to confirm the therapeutic equivalence of two formulations. Those two formulations could be from the same manufacturer in an NDA. They could be different, scaled-up formulation versus the clinical trials formulation, or it could be two different manufacturers trying to product products which perform in exactly, or close to exactly, the same way. So this is a technique that's used in both new drug approvals as well as in generic drug approvals.

And when I say confirmed therapeutic equivalence, you'll see that a lot of what we do,

which other FDA speakers and other speakers will talk about, is there's a great deal of work that goes in on the manufacturer's and sponsor's part on the dosage form design as well as the FDA's assessment of all those things. A lot of chemistry work, which you'll hear from Dr. Duffy, as well as a lot of other work, before we even get to the point of trying to confirm what we already believe by all those other tests. And that's that the products indeed, when and if they are approved, are going to be therapeutically equivalent.

Therapeutically equivalent products, we contend, can be substituted for each other without any adjustment in dose or other additional therapeutic monitoring. And as you see, that's one of the controversial points that was brought up by the previous speaker, and will be addressed at some length later. But that's our contention, when we give an AB rating, that no additional monitoring is required. And that doesn't mean you're not doing the same monitoring you always would do with a patient, but you don't really -- our contention is you don't really need anything extra, any re-titration or so forth. And as you heard, that is one of the controversial points.

And the most efficient method of

confirming therapeutic equivalence is to assure the formulations perform in an equivalent manner. It's a very important concept, and it's something that a lot of the people that I go out and talk to with a variety of different training, pharmacists, physicians, the public, and unfortunately a lot of my FDA colleagues that I talk to as well forget that the bioequivalence we're talking about is actually, strictly speaking, a test of two or perhaps more formulations and how they perform in vivo. And when I say perform, I mean how do they release the drug substance that they contain and make it available for absorption into the body. I mean, that's entirely what we're talking about, and a lot of other clinical concerns that go beyond that are extremely important, but the question, the specific question that we're addressing with this, is are these two formulations, whether it be by the same manufacturer or by different manufacturers, are they going to perform and be equally, or close to equally, bioavailable when I give them under similar conditions to the same patient, or to the same subject. So that's what we're really after with this.

Just to give you a few -- since I'm an FDA speaker I have to quote the regs occasionally. For us, this is a very important -- this isn't just to

quote the regs. This is actually a very important guiding principle for us. Normally the regulations a lot of times are hard to understand, or they're not, you know, not well-written so that normal people can understand it. However, this particular part, which is very important to us who do bioequivalence, is actually very clear-cut, and very based on sound science, and probably sound practice over a good 30 years or so. It lists in this section the methods, the general methods of determining or confirming bioequivalence. And furthermore, it's important to see that this list is not just put up in a random fashion. This is put up in what the writers of these regulations, the scientists who had input into it and the physicians, that it is in order of actual preference, from best and most efficient to least efficient. All of these are effective measurements, used properly, but some are better than others. For oral products whose effects are mediated through systemic effects, which are a great deal of the products that we deal with, the best way to determine whether two formulations release their active drug to the body in the same way are in vivo measurement of that active moiety, or moieties in the biological fluid. And that could be blood or blood plasma. In

the old days they actually measured urine. We don't really do that very much except for one or two specialized dosage forms, or specialized drugs. And so this has proven over a good 30 years with quite a few studies to be the most efficient way at the end. And the end is that very simple thing that I stated, do those two formulations perform in vivo in the same way. So this is virtually all -- every experience that I've ever had with any drug, including the somewhat more complex drugs like this one, this is always the best approach. Now, we may argue what the criteria should be, or whether it should be tighter or looser. But the most efficient means to the end is generally to measure the drug as it appears, first appears in the body and is transported to its site of activity.

Other effects which we have used, and have to use in certain types of products or drugs. We can use in vivo pharmacodynamic comparisons, which is one of the proposals that's being made today. TSH could be considered to fall in that category. Again, we use that for some topical drugs, topical corticosteroids, we use some pharmacodynamic measures for that. It's much more challenging to do that, and required a great deal of effort to get to a point where we could even

do it in a reliable and convincing manner. In vivo limited clinical comparisons. We often don't have a pharmacodynamic measure which can be readily measured, so we actually have to use the same clinical evaluations that were used to approve the drug in the NDA initially, and use patients, and look at the patients' response over time to that therapy. So that is a possibility as well. That's very difficult and challenging to do, clinical responses in general are very variable, you need a lot of patients. At the end sometimes you've done a very large trial and unfortunately, as some of the drug sponsors in the audience will know, you end up with this large effort and not having either a confirmation of bioequivalence or information that says that you've made the wrong formulation and you ought to go back. So you end up with a very equivocal result after putting a lot of patients through a trial. But this does work. If you try hard enough, if you do enough trials, you can get one that either demonstrates bioequivalence or gives you an answer that you haven't made the right formulation and you ought to go back and do it again.

Finally, in vitro comparisons in specific cases, say for -- we have a few non-absorbable GI drugs, and we need to do in vitro comparisons because

you can neither measure the drug in plasma nor can you actually get a very good handle on the clinical effects. Sucralfate is one that's very difficult. That's done with clinical comparisons. Other things like cholestyramine, which binds bile acids in the GI tract we do in vitro binding instead of an in vivo study, and that's proven to be very effective in differentiating like to unlike products. And then the regulations give us, you know, allow us to be creative. When none of the above works, it allows us to go back to science and to actually develop a new method that doesn't even fit in any of the above categories.

This is a slide which I've shown quite a lot. I have two versions. This is the general version for oral drug performance. And the important parts of this -- there are several -- is it lays out in a schematic formulations the steps where you go from a solid oral dosage form all the way to the end to a therapeutic effect. And by therapeutic effect, I include all therapeutic effects, both the desired and the undesired effects, and also pharmacodynamic effects as well. Important point number one is that what we're talking about as far as formulation performance occurs in this step here, in the

transition from that solid tablet to a drug in solution in the GI tract. So the tablet has to disintegrate, and then the particles of drug have to dissolve and become a solution prior to absorption. If the drug is already in solution, then this step really doesn't exist, and virtually all solutions, as far as our regulations and how we handle them, most of the time we don't even do or require in vivo studies, bioequivalence studies on solution dosage forms, unless they have some kind of odd or strange excipient that may affect the absorption. But the vast majority are waived, we don't do any in vivo studies on them at all.

But this point here is the most important point, because that's what the manufacturer puts together, that's what controls how much drug is absorbed and how fast. And so that's really what we're trying to test here. That's the thing that's going to make the difference down the road, if this first step does not -- if the two products do not perform well, or equally, this will lead all the way along to eventually different therapeutic effects.

The other thing that people, especially when I speak to clinicians say is well, you know, you've said you measure blood here, but I'm really

interested in the clinical effects. So why don't you just cut to the chase, cut to the end, and look at the clinical effects, because basically that's what I use in my practice, that's what you used in the clinical trials that showed efficacy, why don't you just measure them directly. It's a very logical comment, but there are some technical problems, I could call them, and characteristics that make this much more difficult to do. And not only difficult as a matter of effort, but difficult meaning that the results I always get are not really definitive when I finally do this trial. The blood concentrations have a fairly linear response. They aren't all that sensitive to the dose that you pick your study to do at, so that the response, meaning the plasma concentrations, tend to be rated in a linear fashion. So it's not exactly sensitive to dose.

Just quickly, this is a much more accurate schematic for levothyroxine or any endogenous hormone where the body stores or produces the drug, and through a feedback mechanism it adds -- the body itself adds more of the same drug or same substance to the blood. So it becomes a little bit more complicated to do blood sampling, since we're already dealing with an endogenous level that we must somehow

subtract out to see what the contribution of the dosage form is. So it's a little bit more complex with levothyroxine or other hormones than the simple case that I just stated.

I have another -- as you work your way from left to right on that scheme, the variability of all those steps goes up, so that by the time you get to clinical responses, you're dealing with quite variable responses, since all of that additive variability. And that's very hard to deal with in studies. It requires large trials.

The other thing about clinical or pharmacodynamic responses is they don't have a linear relationship with their response. There is a part of this curve where I've given a very small amount of drug, and I get no discernible response. There's a portion up here where I've given a lot and I've pretty much maxed out the response that I'm given. If I do my trial up here, I can have a large difference in the delivered dose, and I can see absolutely no difference between the two dosage forms, whereas if I do it on the steep part, which is what is necessary, I can see a very nice sensitivity to differences in dosage form. But it's very, depending on where you are in this curve, that can change, and each person has a

different dose response, each person in your trial.

Just quickly, study designs. We do a two-way crossover, fasted study, and usually a two-way crossover, fed study. There's some alternate dosage -- or alternate study designs here, and those sometimes can be used for specific drugs. Usually with levothyroxine we use the top two, although a suitable alternative properly done, you could do a parallel fasted trial since levothyroxine has rather a long half-life.

And the final, the statistical methods which are always difficult to explain, and since I've pretty much run out of time I won't go into detail about that, but when you hear others refer to AUC and Cmax those are the two pharmacokinetic parameters that represent the extent, or how much is absorbed. So when we compare AUCs from two products we're looking at the entire extent that's absorbed. And the Cmax is related to the rate, how fast it comes in. And so we compare those as well. The data is log transformed. We do an analysis of variance procedure, the statistical procedure with that model that I stated, and from that we calculate those infamous 90 percent confidence intervals that you have heard about. And they must be between 80 to 125 percent.

So as a summary, the bioequivalence is the confirmation of the comparative performance of formulations. And by that we mean the release of the drug substance from the drug product by rate and extent. And this is the final, I guess, thing to understand, that I said we're talking about formulation performance here. Do the two formulations perform in vivo in the same way or not? And that's what we're trying to get at. And there are a lot of other clinical concerns which are important for patient management, but aren't necessarily relevant to this specific and very limited question. And for more information on this I've listed a couple of FDA websites and things which you can look at.

DR. ORLOFF: Thank you, Dr. Conner. Our next speaker is Dr. Eric Duffy. He is a supervisory chemist in the Office of New Drug Chemistry at the Center for Drug Evaluation and Research. He'll be speaking on manufacturing standards for levothyroxine sodium drug products. Dr. Duffy?

DR. DUFFY: Thank you, David. Good morning, everyone. Can I be heard? All right. I just want to take a few moments to discuss some basics. If you're going to study a drug, you need to manufacture it. And at FDA, we spend a considerable

amount of effort to ensure that drug products are manufactured at the highest quality. So I'd like to just -- let's see. I'm going to just briefly describe the drug products, and formulation, and manufacturing basics. And I'll go into a little bit of history about these products. As was indicated, they had been manufactured for a half a century, and most of the time under basically unregulated circumstances. And then the regulatory history as the products evolved, and what the current status is of these drug products.

As was mentioned earlier, the active principle of this drug is an endogenous substance, levothyroxine, which is shorthand designated as T4 quite frequently. It should be noted, and it was indicated earlier, that it has a significant half-life. The half-life is approximately seven days, and that's an important point to note. These products are manufactured as immediate-release tablets. And just to describe very briefly how you manufacture a product, these are -- and I'm sure everyone's familiar with the products being relatively low dose. Very small amount of active ingredient. The active ingredient is blended with inactive components that permits you to actually manufacture a tablet. That's called direct compression. A powder blend is made

which is then fed into a machine that punches a tablet out. And these products are manufactured in batches of millions of tablets. So this is a rather large-scale operation where you have big, huge vats that blend these materials together. One attempts to get a very consistent blend so that tablet after tablet as they're punched in the tablet machine come out in consistent doses. And that's referred to as content uniformity. And this is a very important characteristic of any drug product, but it's most particularly important for a very low-dose drug product. And so the blending process is very important.

Now, these products are manufactured currently under what is referred to as Good Manufacturing Practices. And this is a set of regulations that FDA has which basically codifies manufacturing principles that, if adhered to, result in a high-quality product. And we have -- I work out of Headquarters, but we have people out in the field who actually visit the plants and ensure that the drug products are manufactured under Good Manufacturing Practices.

A brief history of these products. Levothyroxine was first marketed in the 1950s, and as

I mentioned, under non-FDA regulated conditions, circumstances, until 2001. This is a challenging product to manufacture. Levothyroxine itself is relatively unstable, chemically unstable. So one needs to develop a formulation that is designed to enhance its stability so that it can have a reasonably lengthy shelf life for marketing purposes. So it's very important to ensure that one designs a formulation that ensures that the product is stable throughout its shelf life, and retains its potency.

It had been noted earlier by Dr. Galson that FDA had a large number of reports that there was inconsistency in potency across different products and from batch to batch. And this was confirmed in our laboratories that there was indeed a good bit of inconsistency among these products. The products were not necessarily manufactured to try to design 100 percent of the labeled claim. Oftentimes the products were formulated with an excess of the active component so that upon degradation one would still have reasonably close to the label claim amount of drug. And the products did degrade. And I'll show you some data about that later.

Some of the products actually degraded up to something around 20 percent, and that's really

quite significant. When the active ingredient degrades, well it turns into something that's called a degradant, an impurity. And these were not monitored as well. Monitoring of the stability was an important thing. However, the practices across the industry were inconsistent, and were not really according to standards that we currently endorse. So the overall result was relatively inconsistent quality.

As I mentioned, there was not only inconsistency between manufacturers' products from product to product, there was also inconsistency batch to batch within the same manufacturer. The result of that was that some potencies, some strengths, could actually overlap. For example, the super-potent 100 microgram tablet could contain more of the active component than the 112 microgram. And this picture describes essentially what I'm talking about in terms of overlap of dosage strength. If one has something at the high end, for example here, for the 88 microgram tablet, it actually overlaps with the 100 microgram tablet. And so, the prescribing physician doesn't know exactly what dosage strength, when they titrate to dose, they don't know exactly what strength to continue to provide.

Now, after having seen this, observed this

problem in the marketplace, FDA moved to bring these products under our system of regulation. And we issued a number of Federal Register notices, which informed the industry of our intent to bring it under the regulatory umbrella, and these are the citations. We followed up with a guidance to industry about how we were going to proceed with bringing that process under FDA regulation. And that involved a phase-out of unregulated products and a phase-in of the regulated products, which we're attempting to ensure the high-quality standards for.

As Dr. Galson mentioned, we have approved seven applications for levothyroxine products. And as far as I understand, there are four currently marketed in the U.S. In submission of these applications, applications received after August of 2001 were reviewed as generic applications. It should be noted, however, that the chemistry and manufacturing standards are exactly the same whether it's regulated as a new drug application or an abbreviated new drug application -- as a generic application. And I know that quite well because I spent a number of years myself in the Office of Generic Drugs.

Now, the products that we reviewed, the seven applications that we reviewed, are currently

required to be manufactured targeting 100 percent of the label claim at the time of release of the product. And also, to ensure tablet to tablet consistency, the content uniformity also is targeted at 100 percent, although there is some allowable variation, but relatively tight in terms of that variability. The products were required to demonstrate their stability at defined conditions. And this acronym here is International Conference on Harmonization, which is an international agreement, really, of what constitutes appropriate test conditions to demonstrate stability. So products are placed under defined conditions, and the potency and other attributes, dissolution, disintegration, for example, are observed, to ensure that the product retains its specified product quality throughout a certain defined period of time, which was referred to as its expiry, or its shelf life. So these test data are provided to FDA, and we do a suitable analysis of the data to observe the trend toward loss of potency. And based upon these data, we determine an expiry, and agree with the manufacturer on what that expiry should be.

I mentioned that the standards are the same whether they be generic or new drugs. We have a number of manufacturers of drug products. However,

each drug product can be manufactured with an active ingredient provided by some other manufacturer. And that is the most common practice. The active ingredient quality standards are also very important, not only the drug product performance standards, but you have to start with an active ingredient that you know is of a high quality. And so those manufacturers' practices are also scrutinized by FDA, and we ensure that those manufacturers produce a very high quality product for subsequent use by the drug product manufacturer in formulation.

One needs to establish suitable standards for the quality attributes of a drug product. And previous to the regulated approach to these products, the standards were varied widely between manufacturers. There were inconsistent basic specifications. And so we moved to ensure that these standards were made relatively uniform across all manufacturers so that the high quality would be ensured.

I mentioned earlier that we wanted to target at 100 percent of the label claim. And that required some manufacturers to actually reformulate their products to ensure adequate stability of that formulation. The quality standards are now codified

in a monograph in the USP, U.S. Pharmacopeia standards. And there are established, defined dissolution methods, and there are alternatives. There are basically three methods described.

The first point here is with respect to potency. We need to ensure that the potency determinations were done by current state-of-the-art techniques, and that's referred to as HPLC. It's a chromatographic means of determining purity. You'll see there that I've noted that the specification is 90 - 110 percent. Now, that variability is really quite standard across most products. And that is primarily due to simply instrumentation variability, test methodology variability, and a little bit of manufacturing variance. But it's mostly an analytical issue.

Content uniformity, tablet-to-tablet consistency and potency is defined also in the USP under a specific chapter. And in fact, most of the products we have approved have tighter standards than the USP establishes. We also move toward having the impurities, the degradation products monitored to ensure that there weren't any potential safety issues that might result from degradation. And other attributes that are also important for product

performance, such as the tablet hardness, the moisture content which can impact stability, and friability, which means does the tablet break and chip and fall apart. It maintains its integrity. So all these standards were established for each product.

This describes basically what the content uniformity looks like, centered around 100 percent. And there is some degree of variability established. So this is simulated data to show what is typical for a product such as this.

Stability was clearly defined in these applications, and the standards were established based upon the International Conference on Harmonization standards. And also, not only the test conditions are described, but also the frequency of testing to ensure that a suitable amount of data over time is gathered to ensure that you have adequate knowledge of the stability of the product.

Stability of levothyroxine products before we approved the applications was really problematic. And this is also simulated data which just -- it's typical of what we had observed, and how some of the products performed. The blue curve shows products pre-'97, and particularly in the early part of the graph you can see significant degradation, loss of

potency. Products were typically formulated at higher than 100 percent to accommodate this loss of potency over time. Reformulated products shown in the pink -- I hope you can see it up there -- in the pink show that these reformulated products exhibited much better stability performance over time. Starting out with 100 percent label claim, they typically lost just a few percentage points in potency over time. This shows the early part of the curve, demonstrating the dramatic drop in potency for the older products, and relatively good stability being demonstrated with these reformulated products.

And that really concludes my talk on manufacturing. The emphasis I'd like to leave you with is that we have a high degree of confidence that the products that are currently in the marketplace, those approved and in the marketplace, are of high quality, and ensure that the patient receives the proper dose over time from batch to batch, from manufacturer to manufacturer. We have a clear understanding of the quality standards, and we believe that the manufacturers also understand their process and their product, and perform the manufacturing in a manner that produces a high-quality product. Thank you very much for your attention.

DR. ORLOFF: Thank you, Dr. Duffy. Our next speaker is Dr. Henry Malinowski. He's from the Office of Clinical Pharmacology and Biopharmaceutics at the Center for Drug Evaluation and Research. And he's going to speak about bioavailability and bioequivalence studies in the evaluation of new levothyroxine products. Dr. Malinowski?

DR. MALINOWSKI: Thank you, David. Good morning everyone. What I'll be focusing on is the period going from when there were no approved levothyroxine products to the time when NDAs began to be approved. And I'll put particular emphasis on what was done, and why the various steps were undertaken. I would like to emphasize that the issues were not related to the direct safety and efficacy of levothyroxine, the issues were not related to the diagnosis and treatment of thyroid disease, but the issues were much more related to the doubts about the quality and consistency of the marketed levothyroxine products. And that is what FDA addressed by the process which I will be describing.

So what we're trying to say is if a patient is prescribed a 100 microgram dose of levothyroxine, and that's what the tablet says it contains, that it in fact contains as close as

possible to 100 micrograms, that amount of drug. And when the patient swallows this drug, that that drug is released and is made available as close as possible to 100 micrograms of levothyroxine. And then that drug is available for absorption in an efficient and reproducible way. This is what I think has been accomplished by the NDA approval process, and I'll present data to show why I think that this is so.

It has been mentioned, and this describes the issues, these products have been in the market since the 1950s, and none had been approved as a new drug by FDA. There were at least manufacturers and re-packagers out there, and there were numerous reports of therapeutic failures, problems with these products. Related to this FDA took action, and in a Federal Register notice essentially declared levothyroxine a new drug, and indicated that if you want to continue marketing a levothyroxine product, you're going to have to get an NDA approved. And that was done.

Related to that announcement, and this is what I'll be talking about, was an FDA guidance which described what you had to do in order to get an NDA approved. In particular were the bioavailability studies that were necessitated, including a single-

dose (relative) bioavailability study compared to a solution. This was necessary because there was no reference product. So in those cases we use a solution as a reference product. We compare all the products to solution. And also what is called the dosage form proportionality study was conducted involving three different strengths of each product intended for NDA approval. Also, in vitro dissolution testing and so forth was required as part of the NDA approval process.

This is what I see as what the questions were at the time. And they were: Is the bioavailability of the product known? No. Is the bioavailability optimal? That was unknown since we had no idea what the bioavailability of these products was. Do levothyroxine tablets have a proper labeled amount of drug? No. From various literature reports and other sources we knew that this wasn't true. Do the tablets contain a consistent amount of drug? No, again from available information. Does the drug dissolve rapidly and completely? This was unknown. We hadn't seen that data. Is the drug stable over time? No. We knew from numerous reports that this was not the case. I've seen a literature article where an assay was done on one of the products, and it

assayed at 30 percent of the labeled amount of drug. Will subsequent batches perform the same as a batch tested for bioavailability? This was unknown. So these were the questions that needed to be addressed initially as part of the NDA approval process.

Some facts about product stability. Levothyroxine degrades quickly with exposure to light, moisture, oxygen, carbohydrate excipients, and there were numerous recalls, millions and millions of tablets recalled due to content uniformity and other stability-related failures. From the literature I have some information here indicating that up to 109 percent was a starting amount due to the stability concerns. And from this you can imagine how there could be a lot of variation going from even Batch 1 to Batch 2, or Product 1 to Product 2 about how much was actually in the tablet that was being administered.

This is some information from the levothyroxine label. And interestingly, it says that absorption is 40 to 80 percent. Which is it? And 80 percent is actually quite high, and actually the answer is both. And absorption is decreased for levothyroxine quite easily if you take it with soybean, fiber, walnuts, many foods in drugs all decrease the bioavailability of levothyroxine.

However, the 80 percent indicates that levothyroxine can be well absorbed. And that's why the label says `Take on an empty stomach one-half to one hour before breakfast.' I think it's very important that patients be aware of this, and know that you should, for optimal absorption, take levothyroxine tablets with a glass of water, and a period of time before you eat, if it's morning then breakfast, and so forth. Because food, anything you take along with levothyroxine likely will affect its bioavailability getting you closer to that 40 percent number than 80 percent.

Next a little bit about drug absorption and what happens when a patient swallows a tablet, a levothyroxine tablet. In this case, first we get GI transit to the site of absorption. For levothyroxine there is no narrow site of absorption. It can be very well absorbed once it's in solution. After the dosage form travels to a site of absorption there is dissolution of the drug, and then the drug can be absorbed. And I'm showing this diagrammatically here. Starting with the solid dosage form, which disintegrates into granules, which de-aggregates into fine particles. From each of these sources we get dissolution. Primarily, however, the smaller the particles, the faster you're going to get the drug

released and dissolved. And then this results in drug in solution, which can be absorbed. And what I want to focus on is this portion down here. Once we have drug in solution to drug being absorbed. Keep in mind that levothyroxine can be well absorbed if it's just taken with a glass of water. So our goal is to get it in solution. Once we get the drug in solution, any formulation-related factors are gone. We're dealing only with a solution at that point. And levothyroxine, at that point there's nothing complicated about levothyroxine absorption. It's not highly metabolized. It's not actively absorbed. Get it in solution, it can be well absorbed.

How can we validate that this is in fact true? Well, we can validate that by doing -- the first of the two types of studies that I suggested were required for NDA approval. And that is compare a levothyroxine tablet to a levothyroxine solution. And what I've shown here is typical results for that kind of study. And what you see is for the solution, which is slightly higher here, and a tablet of levothyroxine, very similar plasma concentrations. So rapid absorption, complete absorption, and similar absorption to a solution. We saw this again and again in every NDA that was submitted for approval. This is

just a table that shows the same data, just to point out that the Cmax value was even closer, if you look at individual Cmax's which is what are averaged in this table, 14.5, and 15. And essentially identical area under the curve values. So we saw this type of data again and again that levothyroxine can be very well absorbed, similar to a solution. No formulation factors for solutions to be absorbed.

A second study was required for NDA approval also, and I actually see this is as not essential -- it's certainly, it's not essential now for ANDAs. And it was an excellent idea at the time because we knew so little about the products. So what was actually done, and this turns out to be very useful, is that three different strengths of a product were tested. 50 microgram, 100 microgram, and 300 microgram tablets were compared, all at a 600 microgram dose to show -- and what this was important in showing that a manufacturer could make three different batches of a product, and compare their bioavailability. And again, time and time again, as we saw this study in NDAs, we saw this kind of data virtually super-imposable plasma concentration curves similar to the solution study, rapid absorption, and similar absorption for the three strengths that were

tested. Again, these are the data for that table, and the mean comparisons are down here showing how close the Cmax and AUC values were for these products.

So between 1999 and 2000, a number of sponsors submitted NDAs, and the first was approved in August 2000. And there are currently seven approved NDAs for levothyroxine tablets. All of them did the studies that I just described and showed similar results. In addition, other important steps as part of the NDA approval process is sponsors must now target 100 percent of label claim, no unaccountable or stability overages. The days of 109 percent are gone. There is no product on the market that has 109 percent as a starting point, or 105 percent as a starting point. It's 100 percent is the starting point. And that is a major accomplishment. This was a major problem, prior to the NDAs being approved, of differing actual doses among batches and products based on these large overages. In addition, the currently approved products have precise chemistry and manufacturing control requirements, dissolve rapidly, and are stable. Therefore, there are minimal bioavailability concerns. These essentially behave like a solution.

And that rapid dissolution is very

important. We, as part of the NDA approval process established, I believe the number is correct, four separate dissolution tests for the various NDA products. So we did not just set one dissolution test for all of the products. We looked at the data, and companies had to justify using surfactants. If they didn't need surfactants we had them remove surfactants, or lower the amount of surfactants. We set specific specifications for each product, and the seven products were lumped into four different categories. I think there are times when there's too much emphasis placed only on the pivotal bioequivalence study, or the initial bioequivalence study. Patients don't take those tablets. Subsequent to that, companies manufacture another lot, another lot, another lot, and that's what patients take. It is important that companies manufacture the product the same way for each of those batches, and the dissolution test is one of the most important tests, particularly for levothyroxine. If you see the dissolution results for a new batch of levothyroxine, you can relate that to the expected bioavailability for that particular line.

So going back to the questions that were there. Hopefully from what I've presented we can

think of the answers at this point. Is the bioavailability of each of these products in the NDA known? Yes. Is the bioavailability optimal? Yes. Do levothyroxine tablets have a proper labeled amount of drug? Yes. Do the tablets contain a consistent amount of drug? Yes. Does the drug dissolve rapidly and completely? Yes, including specific dissolution tests for individual products. Is the drug stable over time? Yes, that is clearly defined now. Will subsequent batches perform the same as a batch tested for bioavailability? Yes, it's just what I referred to as far as the dissolution testing requirements, the CMC requirements, which are very important for subsequent batches that are manufactured.

So to conclude, the process used by FDA for the seven approved NDAs for levothyroxine products has addressed concerns related to the quality of these products. And I will state that these products can be used with confidence, knowing that the bioavailability and product quality are consistent and high. And any products that fail any of their specifications, assay, content uniformity, dissolution tests, and so forth, will be removed from the market. Thank you.

DR. ORLOFF: Thank you, Hank. Our next speaker is Dr. Barbara Davit. She's from the Office

of Generic Drugs, from the Office of Pharmaceutical Sciences at the Center for Drug Evaluation and Research. And she'll be speaking on Report of Recently Approved Products Performance in Bioequivalence Testing. Dr. Davit?

DR. DAVIT: Good morning. Well, this morning we've previously heard Dr. Conner discuss basic study design and rationale for conducting bioequivalence studies. We've heard Dr. Duffy talk about chemistry manufacturing and controls of levothyroxine sodium tablet products. And Dr. Malinowski has discussed criteria for approval of NDAs, with a focus on bioavailability studies for these levothyroxine sodium tablet products. The objective of my presentation is to discuss those levothyroxine sodium tablet products for which bioequivalence studies have been performed. In other words, submissions for which two levothyroxine sodium tablet products were compared to each other, resulting in a conclusion that the two products were bioequivalent.

First, I'll be talking about the approved levothyroxine sodium tablet products for which these bioequivalence studies were done. Second, I'm going to discuss how the bioequivalence was determined for

these products. In other words, I'm going to discuss the study design that all of these products, all of the applicants submitting NDAs and ANDAs for these products were required to do. I'll present some in vivo and in vitro data from these bioequivalence studies, and I'll finish with a summary and conclusions.

These are the approved levothyroxine sodium tablet products for which bioequivalence studies were conducted. In other words, the two products were compared to each other in bioequivalence submissions. Because all of these bioequivalence studies were successful or acceptable, the products have subsequently been rated therapeutically equivalent. And as Dr. Conner explained previously, therapeutically equivalent products can be substituted for each other without adjusting the dosage or the regimen.

So these comparisons are Levo-T versus Levoxyl, and a second study for Levo-T comparing it to Synthroid. Mylan also has an approved levothyroxine sodium tablet product for which three comparisons were done. One bioequivalence comparison was against Levoxyl, the second against Synthroid, and the third against Unithroid. And finally, there are two

bioequivalence submissions which were acceptable for Unithroid, one against Levoxyl, and the second against Synthroid.

Now we did find that there were variations in the composition of these levothyroxine sodium tablet products. There was a lot of overlap in the inactive ingredients of each of these products. There are some differences too. All of the inactive ingredients that have been used in these levothyroxine sodium tablet products are very commonly used in formulating immediate-release tablets. And the FDA has a lot of experience with evaluating these inactive ingredients. In our experience, we have never seen that any of these inactive ingredients that have been used in these levothyroxine sodium tablet products have affected bioavailability. And as expected, the differences in these inactive ingredients had no effect on the bioavailability or bioequivalence of these levothyroxine sodium tablet products, since all of them did have acceptable bioequivalence studies.

Dr. Conner explained this process briefly earlier, and I'll explain it again. For levothyroxine sodium tablet products, the way in which we determine if the products are bioequivalent to each other is, first, we ask the applicant to conduct an in vivo

study on the highest strength to be marketed. This is generally the 300 microgram tablet strength. If the study is acceptable, we then ask if the applicant wants to develop an entire product line of the various strengths of levothyroxine sodium tablet products. We ask that the applicant show two additional things. In addition to acceptable bioequivalence on the highest strength, the applicant must also submit acceptable in vitro dissolution data on all the strengths of this product line, and demonstrate that all the strengths of the product line are proportionally similar to each other.

And this graph, this is a typical graph showing dissolution data for an entire product line of particular levothyroxine sodium tablet product. These are the dissolution data, and our reviewers in the Division of Bioequivalence, and also our reviewers in the Office of Clinical Pharmacology and Biopharmaceutics and New Drugs evaluate these dissolution profiles very carefully. It's very important that all of the profiles be similar for the lower tablet strengths to be approved. And in this particular case, this is a very good set of dissolution profiles. All of them are comparable, and these data were very strong in support of a finding of

bioequivalence for all the strengths of this particular product line of levothyroxine sodium.

Now this is the basic study design for levothyroxine sodium tablet products. It may seem on the surface like a very simple design, but in reality a lot of thought went into this particular bioequivalence study design. The objective was, obviously, we want the applicant to be able to demonstrate the two products are bioequivalent, but in addition, we want a method that will provide sensitive, accurate, and reproducible means of determining bioequivalence, and also a reasonably conservative means of determining bioequivalence so that not just any two products can be shown to be bioequivalent to each other.

So the basic study design is a randomized two-way crossover design. And in this particular study design this means that all of the subjects receive both the test and the reference product. Now, the test product would be the new product for which the applicant is seeking approval. The reference product would be the product against which the test product is compared. These are small studies. They generally employ no more than 24 to 36 healthy subjects. And we ask applicants to conduct their

studies with both males and females.

The treatments that everyone receives. We ask applicants to give a single 600 microgram dose to both the test and the reference groups. Now, there's two reasons for the 600 microgram dose. One reason is that generally applicants are seeking approval for the 300 microgram strength as the highest strength. And so 600 micrograms, of course, is a multiple of 300. The second reason is that we found that because of a relatively high endogenous baseline of levothyroxine, or T4, it's necessary to give a dose that will give an optimal signal, or a strong enough signal, above the background, or the noise, of the endogenous levels. And we found that a 600 microgram dose was optimal for this.

The washout period is 35 days. Each subject receives the test and the reference product. Because of the seven-day half-life of levothyroxine, we want to allow an optimum time for removal of -- or clearance of levothyroxine from the plasma. And we found that 35 days is optimal. A general rule of thumb, five half-lives is good for a washout period.

Blood sampling is up to 48 hours. And we found that this was important too. We found that 24 hours wasn't quite enough to capture the extent of the

levothyroxine coming from the tablet absorption. More than that, there was too much contribution of the endogenous background, and it was easier for products to pass. Because levothyroxine from the tablet was making less of a contribution, and endogenous concentrations were making more of a contribution. So we found that a 48-hour sampling time was really optimal to give confidence intervals that would assure us the two products were truly bioequivalent.

The analyte that we ask applicants to measure is levothyroxine, or T4. And as Dr. Conner mentioned earlier, the FDA believes that the most sensitive, accurate, and reproducible means of determining bioequivalence is to measure the concentration of the active moiety released from the dosage form in the bloodstream. And in this case, it's levothyroxine.

We ask all applicants to baseline correct, and this has been asked of all the applicants that have submitted acceptable bioequivalence studies without exception. So all the data that I will be presenting later is from bioequivalence studies in which the baseline correction was performed. The bioequivalence metrics on which we ask applicants to perform statistics are the area under the plasma

concentration curve from Time Zero until the end of the 48-hour sampling period, and Cmax. AUC, as Dr. Conner explained earlier, is used as an index of the extent of levothyroxine sodium absorption, and Cmax is used as an index of the rate of product absorption.

And this figure here shows how we determine AUC and Cmax. Cmax is the highest plasma concentration observed visually for each plasma profile. The area under the plasma concentration curve, we have a very simple way of calculating this, and this is by the trapezoidal rule. In other words, we take this plasma concentration profile, divide it into trapezoids, and sum the trapezoids. And we believe that this is the most simple and accurate way of calculating AUC. And before performing the bioequivalence statistics, the baseline is subtracted from the AUC, and as I mentioned earlier, this is required of all the applicants. And for levothyroxine, the baseline actually makes a fairly high contribution to the plasma concentration profile. So a good chunk of the AUC, the non-corrected AUC, is being subtracted. And this really provides an extra level of assurance that the two products are bioequivalent, because this is a very conservative approach. In other words, it can be easier for two

products that are not bioequivalent to pass without baseline correction, whereas if two products are not bioequivalent, there's a much higher likelihood that this is going to be detected with the baseline correction.

Now, there's two bioequivalence statistics that I will present for data. And that's the 90 percent confidence interval and the point estimate. The 90 percent confidence interval is determined using all the geometric mean area under the curve, and Cmax test-to-reference ratios in the bioequivalence study. The point estimate, that's obtained very simply. The geometric means for AUC and Cmax for the test and reference treatments are calculated, and then we take the ratio. And that's the point estimate.

Now this particular schematic shows possible bioequivalence results for a 90 percent confidence interval. Now, the top bar is representative of an acceptable bioequivalence study. And when we say that the 90 percent confidence interval must pass our bioequivalence goalpost, recall, as Dr. Conner mentioned, our bioequivalence goalposts are from 80 to 125 percent. This entire confidence interval must be contained within these limits for a bioequivalence study to be considered

acceptable. And the bell-shaped curve superimposed on top of the top bar is used to illustrate that this represents the population of geometric mean ratios, which we are estimating for these two products based on all the AUC and Cmax ratios that we obtained in the bioequivalence study for both the test levothyroxine product and whatever reference levothyroxine sodium product was used.

Now the second bar shows a failed bioequivalence study. This illustrates how it's possible for two products, the second bar illustrates that it's possible for two products to have a point estimate close to 1, close to 100 percent, but still not pass our bioequivalence criteria. And the reason for this is that the 90 percent confidence interval in this particular case is outside of our 80 to 125 percent goalpost, or bioequivalence limits. So in other words, for a showing of bioequivalence, or a demonstration of bioequivalence, it's not enough that the point estimate be centered on 1 or near 1, the entire confidence interval must fall within these limits.

Now, the lower three bars also show examples of failed bioequivalence studies. If I could call attention to the third bar, this illustrates a

case where the point estimate is relatively far from 1, and as a result, this particular confidence interval falls outside of the bioequivalence limits. And this particular bar shows that it is very difficult, if one is formulating a product, and the mean of the test-to-reference ratios is far from 1, and near either end of the confidence interval, it's very hard for this product to pass our bioequivalence criteria, because it's not enough that the mean ratio fall within the limits. The entire confidence interval must fall within the limits. And the lower two bars just show extremes of products that do not meet our criteria.

Now, keeping this particular figure in mind, the next figure is a graphical depiction of the 90 percent confidence intervals, and the point estimates for the seven bioequivalence studies, or pairs of bioequivalence studies that I presented earlier in the talk. And what this particular figure shows is that the applicants that developed these products were successful in achieving formulations that were bioequivalent to the reference comparators. All of these 90 percent confidence intervals for each of these seven comparisons are well within the FDA's bioequivalence goalposts of 80 to 125 percent.

So in conclusion, several levothyroxine sodium tablet products have been rated therapeutically equivalent to each other. And as expected, variations in the inactive ingredients in these products had no effect on the bioequivalence studies, or the bioavailability of these levothyroxine sodium tablet products. And the FDA has concluded, based on acceptable in vivo bioequivalence studies, and acceptable in vitro bioequivalence data, for each of these seven bioequivalence submissions, that these levothyroxine sodium tablet products are therapeutically equivalent, and therefore substitutable with each other. Thank you very much.

DR. ORLOFF: Thank you, Dr. Davit. Our last speaker in Session 1 is Dr. James Hennessey, associate professor of medicine at the Brown Medical School. He's going to be speaking on limitations of current bioequivalence standards. Dr. Hennessey?

DR. HENNESSEY: Thank you very much. I really appreciate the opportunity to be here, and I absolutely loved all these presentations because it makes it unnecessary for me to try to explain, as is so difficult with clinicians, all this background information. Thank you very much. That was absolutely eloquent.

Well, my job is to try to take what you've just heard and put the vision of a clinician behind it, and how this applies to our patient care, and what our concerns might be with these outcomes. Now, I will also show you a definition of bioequivalence. This is my emphasis and my underlining. I'll read just a bit. It's the absence of a significant difference in the rate and extent to which an active ingredient or active moiety in pharmaceutical equivalence -- no need for me to define that now, good -- becomes available at the site of drug action when administered in the same molar dose under similar conditions in an appropriately designed study, as we've just so elegantly heard described.

Now, from a clinician's point of view, this then talks about the therapeutic effect at the site of activity, which again, from a clinician's point of view is generally measured as a serum TSH, which we utilize to evaluate our patients' therapeutic effect. And so from one definition of bioequivalence, one might conclude that TSH is a useful parameter. Now, especially with drugs that are such narrow therapeutically involved, we've already heard that referred to. And here's a definition from the Code of Federal Regulations that tells us that a narrow

therapeutic ratio drug is one that has less than a twofold difference in the minimum toxic concentrations, and minimum effective concentrations in blood. And as we've already heard referred to, is safe and effective but does require precise titration, as well as patient monitoring.

Now, the data from the Carr Study is a great illustration of why levothyroxine is a narrow therapeutic drug. The Carr Study was done on 21 hypothyroid patients who were studied every six weeks on a series of different levothyroxine doses. Assessments were made of these patients approximately six to eight hours after they ingested their levothyroxine prior to breakfast. And when they came in for their evaluation, they had their pill counts counted so that compliance could be assured. They had clinical parameters measured, such as weight, pulse, Billewicz scores, and a questionnaire of general wellbeing, and had biochemical evaluations with a basal TSH, or free T4, free T3, and then a TSH after TRH stimulation, which at the time was state of the art and the most sensitive way of approaching the hypothalamic-pituitary axis. They were considered to be at an optimal dose of levothyroxine if their TRH-induced TSH response fell within the reference

interval of 4.7 to 25, and were therefore considered to be truly euthyroid by the then state-of-the-art methodology. Then their doses were modified by 25 or 50 micrograms, and they were reevaluated six weeks later.

This shows again what Dr. Ladenson showed us earlier, that at optimum dose, these are the basal TSH values for these patients, and minor decreases in levothyroxine, over here 25 micrograms and over here 50 micrograms, led to considerable increase in the TSH values. Similarly, when the dose was increased by either 25 micrograms, 50 micrograms, or 75 micrograms, the majority of patients became considered clinically thyrotoxic based upon the clinical parameter of TSH that was being utilized. And by the time they were 50 micrograms overdosed, then indeed 100 percent were classified as thyrotoxic. So this study truly shows the narrow therapeutic index in thyroxine, and reinforces the concept that small changes in the thyroxine dose result in changes in our clinical assessment of patients. So as a clinician, I'm going to consider someone thyrotoxic if their TSH is suppressed, or hypothyroid if their TSH is elevated.

Now, in this study, the average dose at optimal was 108 micrograms per day, which makes the 25

micrograms less than optimal, less than a 0.25-fold change, clearly meeting the definition of a narrow therapeutic drug. And this results in 89 percent of these folks at 25 micrograms of being hypothyroid. And of course, that's a majority that's even filibuster-proof. The 25 micrograms more than optimal dose, also a less than 0.25-fold change, results in a 55 percent majority of the patients being classified as thyrotoxic, which of course could be achieved as the majority with cloture.

When we look at what patients and physicians are working with on a daily basis, with the FDA-approved doses that we have to work with, we see in the blue scale here that the differences are less than 25 percent in the majority of the doses that are available. And if we look at the circled values here, we see that several of these doses which are clinically useful, and utilized on a regular basis, range from 9 percent to 12 percent. And those two numbers will come up again. So, very small dosage changes are recognized in clinical practice as having a clinical impact. And indeed, it would be sort of difficult for a clinician to believe that switching from 100 to 112 micrograms would not have any meaning, as well as not being able to have the confidence that

staying on 100 micrograms might not mean that their patient was receiving 112 micrograms.

So the purpose of bioequivalence, as we've heard very elegantly outlined, is to demonstrate that there is indeed therapeutic equivalence. And it is to assure that these products can be substituted without concern for adjustment in drug dosage, or the need for any follow-up in therapeutic monitoring, which I believe we would all agree is our goal. It's been said that the most efficient method for assuring this is to assure that the formulations perform in an equivalent manner. And I believe we're only parting our paths here because we don't necessarily agree on what the manner should be in which the patient should be assessed. As we've already seen in order of preference, the pharmacokinetic studies are on top, and we've already heard justification for that. It's because the measuring of the active ingredient at the site of action per se is not feasible, and therefore measuring the blood levels is the substitute because PK is a bioassay of the absorption of the active ingredient.

So that brings us to this portion of the cascade of events -- and again, I want to thank Dr. Conner for this wonderful slide that I've used on

several occasions now, because it is so clear -- while measuring the blood levels to make an assessment of the comparability of these drugs. The clinical questions that are raised, however, when clinicians think about this issue are `Are these limits of acceptability simply too wide with a narrow therapeutic range medication such as levothyroxine?' Certainly the 90 percent confidence interval falling within 80 to 125 percent acceptance range allows detection of 20 percent differences with great assurance. But what differences are clinically appropriate, and is a 20 percent difference clinically appropriate or potentially not, and what we would like to be able to investigate further is what differences can be detected. So the first step in doing this, I believe, would be to take a look at the now updated PK methods and see how they perform in comparison to the previous PK methods.

So this was done in a study of 36 healthy volunteers directly out the playbook, with an even match of men and women. They underwent fasting, open label, randomized, three-period crossover study. Now here, the washout periods between the study periods was lengthened to evaluate the potential that there might be some carryover with the superphysiologic

doses of thyroxine being administered. These people were treated with specifically three different doses of levothyroxine, all of which came from the same brand and the same lot to assure as much lack of variability in those other aspects of the dissolution solution, so that we could take a look at 600 versus 450 micrograms versus 400 micrograms to see if the pharmacokinetic methods could detect these differences with assurety.

Uncorrected, the 600 microgram versus 400 microgram dose, as well as the 450 versus 600 microgram dose, and the 450 versus 400 microgram dose all appeared to have their 90 percent confidence intervals between 80 and 125 percent. But after correction, the 33 percent difference noted here, as well as the 25 percent difference here, was clearly detected, which obviously we've just been informed, led to the adoption of the baseline correction in the pharmacokinetic methods, which of course is very good. However, there is some concern in the clinical community about this 12.5 percent difference that does not seem to be detected in this particular protocol.

Well, the clinical questions then are asked of me as I discuss this with clinicians around the country are what differences then will this

pharmacokinetic method actually pick up? Would it be the average of about up to 3.5 percent as meta-analyses of previous trials, or assessments, seem to indicate from these two publications that were both out in JAMA? Is it a 9 percent difference, as I think we would all agree we have stated on several occasions would be meaningful in a clinical sense, hence why would we have dose increments that are as small as 9 percent. Is it a 13 percent difference, which is just a little bit higher than the 12.5 percent differences that are seen in the midrange of those things, or is it simply something less than 20 percent. What difference in bioavailability would be acceptable as bioequivalence? Well, this is data from the supplemental NDA application of the Levo-T product being distributed by Sandoz versus Synthroid and Levoxyl. The rules were followed here to a T, and they use 600 microgram doses under fasting conditions with the stipulated 35-day washout, and standard pharmacokinetic parameters were measured.

This is, as you just saw, thank you, the 90 percent confidence interval for the Sandoz versus Synthroid comparison. And this is the Sandoz versus the Levoxyl comparison. Both 90 percent confidence intervals pass the 80 to 125 percent goalposts,

indicating that from a pharmacokinetic viewpoint, these are bioequivalent. From a clinician's point of view, however, we think of it slightly differently. When we look at the Levoxyl comparison over here, we are not particularly impressed with the 2.3 difference in the relative bioavailability between these two products, but much concern has been voiced to me, as people have seen this data, with a 12.5 difference, apparent difference in relative bioavailability in these comparisons with Synthroid and the Levo-T product. More recently, the data from the other comparisons has been put into the public domain, and here we see a slide that is not in your handouts, but reiterates the 12.5 percent difference in the Sandoz versus Synthroid comparison, and look at all of the AB2 rated drugs, AB2 being the drugs that use Synthroid as a reference. And here's the Mylan comparison to Synthroid, with 109 percent relative bioavailability difference, and the Unithroid comparison with 103 percent relative bioavailability comparison. Now, the asterisks affixed to these bars indicates that the 90 percent confidence interval exceeds the 9 percent difference in that 90 percent confidence interval. So, from a clinical point of view, we are seeing 12.5 percent difference, 9 percent

difference, and about 3 percent difference as we go along. And we have concerns, because we know these are doses and dose increments that we make in our patients on a daily basis.

Looking at the AB3 rated drugs to Levoxyl, we see the previously stated Sandoz data here at -2.3 percent, and the 2 percent difference noted for the Mylan comparison, with a 2.7 percent difference noted in the Unithroid comparison. Here, again, the 90 percent confidence interval exceeds the 9 percent difference potential between these two products. So, in conclusion, the clinical community and FDA have advanced precision in clinical monitoring and delivery of high-quality thyroid hormone products for therapy. Each step of this standardization has moved us closer to our goal of achieving consistent, precise levothyroxine preparations to enhance patient care outcomes, and the PK assessment, however, leads to some concern in the clinical community that we may be falling short of assuring that we have true interchangeability of these products, which would be necessary for consistent, precise dosing. Thank you for your attention.

DR. ORLOFF: Thank you, Dr. Hennessey. I think we'll take a 15-minute break at this point, or a

20-minute break, and we'll return at 10 minutes of 11:00 for the public comment period.

(Whereupon, the foregoing matter went off the record at 10:29 a.m. and went back on the record at 10:54 a.m.).

DR. ORLOFF: Okay. Let's get started again. For the next hour, we've devoted the time to four speakers from the regulated industry. The first speaker is Dr. John Leonard, representing Abbott Pharmaceuticals. And he'll speak for approximately 20 minutes.

DR. LEONARD: Thank you. I'm John Leonard, vice president of medical and scientific affairs at Abbott. We appreciate the opportunity to share some of our thoughts with the workshop here today. Abbott's the manufacturer of Synthroid, a widely prescribed levothyroxine product. I come to this workshop as a manufacturer, understanding what it means to produce a product. I also come as a physician who's mindful of the conditions for which these products are used. I'll discuss both perspectives, and describe why we and virtually the entire endocrine treatment community believe that this workshop is not about discussing dry regulatory issues, but instead critically important medical

questions. These are medical questions that should be addressed very carefully before proceeding further down the path that assumes therapeutic equivalence and permits widespread switching of agents that are used in highly individualized therapy, regardless of who manufactures these agents. Let's review why this is so.

Thyroid gland produces LT4 hormone essential to life, and we've heard about that. Because the thyroid produces an essential hormone, the body developed a finely tuned mechanism to assure that thyroid hormone is present in appropriate levels. These levels vary relatively little within a patient day to day. When the thyroid is diseased, this delicate balance is disrupted. Hypothyroidism manifests with well known effects illustrated here, and hyperthyroidism also causes many medical conditions, each highly prevalent.

Well, what's the goal of thyroid hormone replacement therapy? The doctors attempting to replicate the finely tuned homeostatic state that's essential to human health, at best we can only approximate this goal. When a physician initiates thyroid hormone therapy, a titration process is carried out to achieve the appropriate dose. Doctors

provide microgram doses to patients, with dose increments differing by as little as 9 percent, as we've heard. These tiny dose increments are essential to good titration, and are critical to achieving the optimized treatment regimen for each patient. Clinical indicators provide gross indications over improvement, but the titration is further informed by serum TSH levels, the body's internal thermostat for LT4 effects. Ultimately, physicians supplement clinical observation and biochemical tests with a highly discerning indicator of treatment success, asking a patient how he or she feels. Once the patient feels well, great attention is placed on keeping the patient well by minimizing variations to the treatment regimen.

Some degree of variability surrounds any treatment regimen for any medical condition. Minimizing that variability is always desirable, but particularly so when giving LT4. Most drug regimens provide a chemical exogenous to the body, one that is not part of its homeostatic mechanism. Because they are extrinsic to the body, the body is forgiving of major variability. Levothyroxine, in distinction to almost all other medications, is a replicate of an agent that the body itself produces, and is one of the

pillars of the body's homeostatic mechanisms.

Clinical experts emphasize the importance of minimizing variability in LT4 therapy. They recognize that additional variability is introduced by differences in bioavailability across different formulations of LT4. These clinical experts, and the societies that represent the vast majority of endocrinologists urge avoiding any source of variability introduced unnecessarily into the treatment regimens. They identify vulnerable patient populations as being at the highest risk for the consequences of over- or under-treatment. For many, the clinical consequences, when they occur, are profound and not reversible.

The FDA also recognized the importance of minimizing variability in treatment regimens. They required all makers of levothyroxine to submit NDAs. They determined that the NDA process would assure control of manufacturing variability, and that has been achieved, as pointed out already this morning. In 2001, they stated their intention to control refill-to-refill variability to 9 percent or less, then reiterated this target just last year. In July 2004, FDA assured manufacturers and the clinical community that its standards will not allow products

that differ by 9 percent or more in potency or bioavailability to be rated therapeutically equivalent. This target was set to reduce the medical consequences of introducing variability into these products.

The clinical consequences of missing the optimal targeted state are profound from either insufficient or excess LT4. These consequences can present with disastrous medical outcomes. After a child is born is the wrong time to realize that a mother has been under-treated with LT4 during her early pregnancy. The damage is done. Likewise, osteoporosis discovered at the time of hip fracture, or afib discovered at the time of stroke or MI is the wrong time to identify that too much levothyroxine hormone was administered. The damage is done.

What are the sources of variability that doctors must overcome? How do doctors and patients contend with these sources of variability as they chart a course of treatment? They recognize that LT4 variability is additive. Each source of uncertainty in a treatment regimen is an element that must be accounted for and overcome by some strategy.

These sources of variability can be grouped into two categories. The first are

variabilities that we know and manage. These are risks that doctors, patients and manufacturers identified and studied. We have treatment strategies that are usually successful in overcoming these sources of variability. The second category of variability is new and not understood. Strategies to overcome this newly introduced variability have not been devised and tested. We must therefore consider any approach to addressing this new source of variability at best hypothetical, and more strictly unknown.

What are these sources of variability that doctors treating thyroid disorders must overcome? The set of known and managed sources of variability contain two main elements. The first is intra-product variability, and the second consists of human factors. Each is inherent to treating any condition with any product, regardless of the therapeutic intention. But variability in patients receiving LT4 therapy is particularly consequential because LT4 is replacing an endogenous hormone essential to the body's homeostasis, unlike most drugs that are not replacements for hormones made by the body.

Intra-product variability is the first variability that we know and have devised strategies

to manage. FDA took action to ensure that this source of variability was addressed via the manufacturing controls that come with NDAs. Any medication has some inherent chemical variability. It's precisely because of this that all medications, including LT4, carry expiration dating displayed on each batch of product. This dating gives confidence that the variability of that product lies within a known range and is controlled by careful monitoring. Although tight limits surround release specifications for each LT4 product from any given manufacturer, differences of bioavailability across products result in a widening of the total range when all products are considered as a class. This is highly undesirable.

Human factors are the second category of known and managed sources of variability. We know that like any substance presented to the body, the absorption of LT4 can be influenced by food and other drugs. We also know that patient compliance can vary person to person. We address these human factors directly by two important means, both at the level of the doctor and patient. First, doctors engage and influence their patients directly via face-to-face encounters. Many opportunities exist for ongoing counseling to control these factors over time. In

addition to counseling is the titration process by which therapy is individualized. Individualized therapy is fundamental to overcoming the variability in a patient's diet, concomitant medications, and compliance patterns. Because titration is carried out over weeks or months, it is an excellent tool to identify, integrate, and address the variability emanating from the human factors of any individual. This is how we have successfully carried out LT4 replacement therapy for years.

Variability is cumulative. Each additional source of variability in levothyroxine is another hurdle that the physician must overcome while attempting to establish the euthyroid state, or diverse therapeutic target. We have now introduced another source of variability into the treatment of thyroid disorders. It is a source of variability that is new, and strategies to overcome that variability are untested, and therefore their adequacy is unknown. This I believe constitutes a real but unnecessary risk for patients taking LT4 products. This new risk is product-switching based on assumed therapeutic equivalence. While product-switching for most products for which bioequivalence has been established is usually not an issue, it is far from certain that

this applies to LT4.

What is the standard by which product-switching is permitted? When we term products "interchangeable" what do we accept as close enough? When products are deemed interchangeable, it is different from saying that they are identical. Products are deemed interchangeable when they are found to have bioavailability characteristics that lie within a pre-specified statistical range, as we've heard. We use statistical limits to say that products are close enough to each other to be considered interchangeable. The PK characteristics we examine must then have the extent of their variability lie within boundaries that are within 80 to 125 percent of the performance characteristics of the reference product. This is a range used for many products over the years, and it has served us well. However, it is usually a limit used for drugs that are exogenous to the body, and have little to no direct role in maintaining the body's homeostatic state.

A fundamental question is whether this set of boundaries is acceptable for endogenous hormones such as LT4. Can we assume one size fits all? We heard that these boundaries are used, but we did not hear why they should apply to LT4. This question is

fundamental, not so much because it is a regulatory standard laid out years ago and applied to products not produced by the body, but because in the case of LT4 it is a medical question. Have we established that the bioequivalence standards implying therapeutic equivalence for products like Prozac and penicillin apply to hormones the body itself makes? Where is the data showing this? This medical question has been explored only in a cursory fashion. In fact, we now know that, based on clinical testing, the bioavailability standards for LT4 products will lead to the approval of products that are known to vary by 12.5 percent. Is this appropriate for this class of medication?

This variability is not a theoretical concern, it's a reality. Consider the case of four levothyroxine products which we've heard about. We will treat Synthroid as a reference product, and compare relative bioavailability of other products considered seamlessly interchangeable. The bottom axis shows the relative bioavailabilities, but it can also be considered practically a Synthroid microgram dose equivalence. If a dose of Synthroid is found to have relative bioavailability of 1, we record that as such. A recently approved version of levothyroxine

was found to have a relative bioavailability of 1.03 compared to Synthroid, another 1.09, and another 1.125. Around these point estimates there is a range of variability as indicated here.

There is no inherent issue with any one of these agents by themselves because patients will be titrated to their targeted level on an individual basis, so long as patients remain on the agent which they were titrated. But what has not been tested is whether patients can safely move from one product to another. Imagine if a patient were titrated to a 100 microgram dose of Synthroid, and was then switched to the Sandoz product. It is as if the patient is now receiving 112 micrograms of Synthroid instead of the 100 microgram dose for which he was titrated. This is a form of variability that the physician did not anticipate, and thus did not address via titration. It is a form of variability introduced unbeknownst to the doctor. When this much variation is allowed for a hormone, what is a doctor to do? Should he read each product's NDA and ANDA to compensate? As you can see, we've traded the intra-product concerns discussed earlier for uncontrolled inter-product concerns.

Well, what might be the consequences when many patients are switched from the agent on which

they were initially titrated? This analysis illustrates such an example. A simulated population of 200 patients is titrated to TSH levels between 0.4 and 4 typically targets. Note that when TSH levels fall due to high LT4 levels, a hyperthyroid state is achieved as denoted by the red line. There are no abrupt cutoffs, but the likelihood of afib and other manifestations of hyperthyroidism climb as one moves further below the red line. As TSH levels rise due to low LT4, the manifestations of hypothyroidism increase, especially as one moves increasingly beyond the green line. If one introduces a switch of LT4 preparations varying by 12.5 percent, this can happen based on approved products. The population responds to the more bioavailable formulation by reducing the median TSH levels. The median patient lies within the desired TSH boundaries, but half of all the patients lie above this median value, and half lie below it.

It's clear that the median levels do not tell the whole story. We retain the median patient as before, but now we also cull out the most extreme 10 percent of patient TSH levels. Under these conditions, we have taken patients who were within our targeted boundaries at the outset and have pushed them unwittingly into values well outside of our targets.

These patients, if presenting to a medical clinic, will likely have their LT4 doses reduced in response to the low TSH levels. In this case, products were clearly not seamlessly interchangeable. And especially worrisome is that the prescribing physician may not even know that a switch took place after the prescription was written. Remember that in this example we are talking about 1 in 10 patients who switched but become hyperthyroid. And recall that about 13 million Americans take LT4 products.

The prior example is the result of a simulated switch of LT4 and its consequences on TSH levels. Firm epidemiological observations have established the association of depressed TSH levels in afib. In these data, more than 2,000 members of the original Framingham cohort were followed to determine the incidence of afib and its relationship to baseline TSH levels during a 10-year period. The Framingham data indicate that with slightly low levels of TSH, as indicated by the green line, the relative risk of afib over time is about 1.6 relative to people with normal TSH. At lower levels of TSH, the relative risk climbs substantially, with the risk estimated to be 3.1 times that for normal. It is obvious that maintaining TSH levels close to normal is an important public health

objective.

We can apply this information to our test group in which we introduce a simulated switch of products with the relative bioavailability differing by a factor of 1.12. By anticipating the changes to TSH, we expect that for every 1 million patient years of switching, there will be in excess of 1,200 cases of new afib. Just as with afib, one would expect to have additional cases of MI, and other well known consequences of hyperthyroidism.

One question raised by statistics such as these is where are all the projected adverse events? The answer to this question is straightforward. The conditions associated with both hypo- and hyperthyroidism are highly prevalent in the United States. Over two million people have afib in the United States and about 160,000 new cases occur annually. With a background incidence this high, the incremental incidence of afib will easily be overwhelmed by the vast number of cases already present. These thousands of new cases will only be an increase of about 1 to 2 percent in the overall incidence, or less than 1 percent in the overall prevalence. These rates will only be observed by careful observation, but the tools now in place are

unlikely to suffice. Because doctors do not know a switch has occurred, they will not link an AE to the switch. This is also true for the incidences of MI, osteoporosis, and other manifestations of inappropriate LT4 treatment caused by switching.

We all believe that patient health and safety is the paramount goal. But as we pursue that goal, we must confront some questions. Do we really know what variability among products truly allows for seamless interchangeability? What data assure us that criteria applied to standard drugs are equally applicable to this endogenous hormone? Do we really have appropriate tools in our hands to determine the corrected relative bioavailability of these products? As it is, we now do studies in healthy volunteers with impact thyroid glands. This seems like an obvious problem, as the thyroid gland in these healthy volunteers works to minimize variations among test agents by its own powerful homeostatic properties. Do we really understand the relationship of variability to the underlying risks in different patient populations, such as kids, cancer patients, and the elderly with heart disease? Why introduce yet another source of variability into this huge patient population? In a setting in which more than 13

million people, or 1 out of every 19 Americans receives LT4 products, what appear to be small differences become big numbers.

So what have we gained? If we do not really have good tools to determine bioequivalence, if small differences matter, if treatment standards are not well developed to address the newly introduced variability, and if the clinical experts all point to this as a medical issue, this all reduces to a simple question. Is the additional variability introduced from switching LT4 products worth the risk to patients? Thank you.

DR. ORLOFF: Next speaker is Michael Lamson, M.D., from King Pharmaceuticals.

DR. LAMSON: High-grade disease. My name is Mike Lamson. I am an employee of King Pharmaceuticals. We are the makers of Levoxyl.

I would first like to say that King Pharmaceuticals agrees with Abbott's original citizen's petition for reconsideration of T4 guidances. However, we would like to present the results of two bioavailability studies because it is our belief that we can learn a lot about optimal T4 dosing with these guidances, and some of it we feel may be important to the issue of interchangeability.

The first study was a comparative bioavailability study where Levoxyl was compared to what I'm going to call Brand B. I think for purposes of this meeting we want it to be more educational and not a marketing promotional presentation. But I've got approximately nine slides that I'll hope to get through in about nine minutes. In terms of the in vitro characteristics, Levoxyl and Brand B are widely prescribed commercial T4 products. Both meet USP dissolution specifications. And as an FYI, Levoxyl, although it is not classified as an oral dissolving tablet, it is a rapidly dissolving tablet. Basically it approaches 90 percent dissolution within 2.5 minutes. It basically dissolves when it comes in contact with a moist surface.

This first study design made use of the FDA's T4 guidance. It was a randomized open label two-way crossover study in normal volunteers. We also have in our studies increased the number of subjects because we also believe that the acceptance interval, we want that to be as narrow as possible. So we generally run our studies with N's on the order of between 40 and 50 subjects. But these normal volunteers each received a 600 microgram dose under fasted conditions with 240 ml's of water. There was a

35-day washout period between doses, and we corrected for baseline concentrations by subtracting the mean of the initial three values.

Here are the results of the first study. You can see the mean PK parameters in the middle for Levoxyl and Brand B. The pharmacokinetic parameters are shown in the left-hand column. You can see the two -- what have become the primary pharmacokinetic parameters for levothyroxine, and that is Cmax and area under the curve from Time Zero to Tmax, where T is usually 48 hours, but it could be 24, 48, 72 hours, or it could be the last quantifiable concentration. And here are the PK parameters here. Over on the right we see the bioequivalence parameters where we use Brand B as the test product and Levoxyl as the reference for comparison. What we list here is the geometric mean ratio, and the 90 percent confidence interval. As you can see here, the 90 percent confidence interval falls within the acceptance range, and also includes a value of 100 percent. By some standards, I suppose, one could argue that these products are dead-on bioequivalent. However, if we take a look at some of the other PK parameters that are not usually included in bioequivalence assessment, but nonetheless important for bioavailability, in

particular Tmax, you can see there were subtle differences in the rate of absorption that were really not reflected by P concentration, but were reflected by Tmax. The median Tmax for Levoxyl was two hours, the median Tmax for Brand B was three hours. And in fact the averages, I think the average for Levoxyl was about two and one-half hours. The average for the Brand B product was over four hours.

And there are no bioequivalence statistics that can be used to assess these differences. However, Tmax can be used to define something called partial area under the curve, which is a metric that's sometimes used to assess what we call early bioavailability. And this is not something that King invented. Actually, Ni Ling Chang and others, including some of our panelists, have considered partial AUC as an assessment of early bioavailability for a number of products. When it's employed here, partial AUC generally refers to the area under the curve from Time Zero to the median value of the reference product, or sometimes the faster absorbing product. In both cases that was Levoxyl. And as you can see, the area under the curve, or what we call the partial area under the curve, from Time Zero to two hours, here are the mean parameters here and

coefficients of variation. And if we apply the bioequivalence parameters, we can see that the bioavailability for Brand B is 23 percent less than that of Levoxyl, and the 90 percent confidence interval falls well below the acceptance interval. So in a sense, even though these two products have been shown by usual bioequivalence standards to be equivalent, when you consider early bioavailability of T4 products, they're not the same.

Looking at this in a little bit different way, here are the baseline corrected T4 concentrations from Time Zero to 2.5 hours, just to really illustrate the point that what I'm talking about in terms of a 23 percent difference in bioavailability represents this region right here between these two curves.

Is assessment of bioavailability important for T4? Well, at King Pharmaceuticals we think it is, especially when you take into consideration how little we know about food-drug interactions with this particular class of drugs. For example, if you look at the class labeling, we actually have two different recommendations, one for drugs and one for food. For drugs, it says the T4 should be taken at least four hours apart from drugs that interfere with T4 absorption. These include antacids, bile acid

sequestrants, ferrous sulfate, and sucralfate, among a list of many products that can be found on the label. On the other hand, food, it says that T4 should be taken on an empty stomach at least one half hour before a meal. And examples of food interaction include soybean flour, which is a component of infant formula, cottonseed meal, walnuts, and dietary fiber. I don't know how many people have infant formula for breakfast or walnuts, but certainly dietary fiber would be a consideration. But it makes you wonder. Much of this is not so much related to diminishing the dissolution characteristics of the drug. But these are factors which can, when they come in contact with T4, can bind to it and prevent its absorption. And it makes you wonder why we have two different class labels when we're talking about the same phenomenon, one for drugs that says four hours, one for food that says one half hour.

Second study I'd like to talk about is a food effects study. And here we made use of two guidances, the T4 guidance for the study design and the food effect guidance for the treatment design. Levoxyl again is greater than 90 percent dissolved in 2.5 minutes. This was a randomized three-way crossover study with 48 subjects who received a single

dose with a 35-day washout period.

The meal consisted of a standard high-fat breakfast, typical FDA breakfast here. It was 950 calories, 16 percent protein, 26 percent carbohydrate, 58 percent fat. I suppose we could be criticized for the way the drug was administered. We administered the drug four hours before a meal -- that represented fasted conditions -- 10 minutes before a meal, and immediately after the meal. We were doing this in isolation, so one thing we couldn't risk, or me personally, is to basically show for one of the fastest releasing products on the market, we're the only ones who couldn't follow the class guidance for food effects. So we in this particular study could not look at the 30-minute period. And some could also argue that we're giving a superphysiologic dose, and we're also probably giving a superphysiologic meal in this particular study.

Here are the results of that study. You can see the T4 concentrations under fasted conditions as represented by the blue line, and the other extreme, the red line represents the T4 concentrations when the drug was administered immediately after the meal, where you see diminished rate of absorption, as well as a substantial reduction in the overall

bioavailability. The more interesting result was when you take this rapidly dissolving tablet and administer it 10 minutes before a meal, there did not appear to be a reduction in the rate of absorption. However, it did become very clear to us that even when the drug is in a solubilized form, when it comes in contact with something like food, there is a significant, actually substantial reduction in bioavailability. And as you can see in this next slide, when we look at the geometric mean ratio, the 90 percent confidence interval, the overall food effect is on the order of about 40 percent, a 40 percent reduction in bioavailability, which is a huge number because an awful lot of our experts at this meeting have been talking about T4 products and interchangeability, and the fact that small adjustments in the dose, or small differences in bioavailability can product logarithmic changes in response, as measured by TSH. And we think that's important.

One of the last few slides here. If we take a closer look at early bioavailability for the food effect study from Time Zero out to two hours we can see here is the profile under fasted conditions, here is what happens when you administer the drug immediately after a meal, and here is what happens

when the drug is taken before the meal. And there's no interaction until out after 0.5 hours. But as you can see here, most of the action occurs between 0.5 and two hours. I think this particular figure highlights the importance of early bioavailability because it is over this period, for Levoxyl anyway, over this zero to two-hour period that T4 has the potential to come in contact with something that could decrease its bioavailability.

And one final slide. I'd just like to say that points to consider in addition to alternative means of equivalence testing. Pharmacologic methods such as AUC should be used to assess early bioavailability. Food effects studies should be conducted to optimize therapy with respect to class labeling, and ask the question is one half hour dosing before a meal long enough for all products. And also we recommend food effects studies should be required of all T4 products for purposes of labeling and establishing interchangeability. We might find that the proximity of dosing in relation to a meal could be one half hour for Product X. It could be one or two hours for Product Y. And even though these products have been shown to be bioequivalent, there might be differences and these products might not be

interchangeable. Thank you.

DR. ORLOFF: The next speaker, Frank Sisto, Mylan Pharmaceuticals.

MR. SISTO: Good morning. My name is Frank Sisto, and I'm the vice president of regulatory affairs for Mylan Pharmaceuticals. I promise to be brief so that -- allow time for my colleagues from Sandoz to complete their presentation.

Mylan Pharmaceuticals has been developing, manufacturing, and marketing generic drug products for a number of years. Mylan is a well known and respected generic drug company, and on behalf of its employees I'd like to say that we take great pride in our ability to manufacture, develop, and market quality bioequivalent generic pharmaceuticals to those in need.

I have been with Mylan almost 10 years, and in that period of time I have been involved in the development, review, submission review and approval of approximately 200 applications for new generic drug products. Mylan has a long history in working with the FDA's bioequivalence requirements. We believe that the FDA criteria for demonstrating the bioequivalence of generic versions of levothyroxine provide acceptable methodologies for establishing such

equivalence. These criteria are considered satisfactory for establishing that the generic product is safe, effective, and therapeutically equivalent to its name-brand counterparts. In addition to these in vivo requirements, a generic drug product must meet other FDA physical and chemical requirements to confirm that it will maintain the quality, strength, and purity that it claims to possess throughout its proposed shelf life.

As you heard Dr. Duffy and Dr. Malinowski this morning, one of the primary issues that caused FDA to take action back in 1997 was the quality and consistency of the products that were currently being marketed at that time. Since the approval of Mylan's generic levothyroxine in June of 2002 through April of this year, we have manufactured a total of 160 lots, covering all 11 product strengths for which we currently have approval. As you can see on this slide, the average assay values for all those 160 lots tested range between 99 to 101 percent of label claim. The mean values for content uniformity of these 160 lots range between 99.9 and 101.6 percent, with relative standard deviations ranging from between 1.4 and 1.8. As you can also see, the average dissolution values for all 160 tested, which have a specification

of not less than 70 percent dissolution in 45 minutes, range from between 83 to 87 percent at the time of manufacture.

And again, while this is important criteria for the release of these products, what is very critical is that these products remain stable throughout their proposed shelf life. The stability history of Mylan's generic levothyroxine product also shows that we have a very stable product with very consistent results. For those product lots that have reached the 24-month stability time point, the average assay value for all lots tested have been between 95.7 and 102.4 percent, demonstrating very minimal loss in potency after two years. And again, looking at the dissolution data with a limit of not less than 70 percent dissolved in 45 minutes, this showed a range of between 81 to 85 percent for those lots tested at 24 months, again demonstrating a very stable product.

To further support the therapeutic equivalence of Mylan's product, I would like to share with you the data that we have collected with regard to adverse events from Mylan's levothyroxine product. Mylan was first approved as an AB rated therapeutically equivalent generic to Jerome Stevens Unithroid in June of 2002. We subsequently attained

approval as a generic equivalent to Jones Pharma's Levoxyl and to Abbott's Synthroid. And we began marketing levothyroxine in December of 2002. From December 2002 until April of this year, we have only had 32 adverse event reports. During this period, there have been over five million prescriptions dispensed with Mylan's levothyroxine product. This equates to 0.006 adverse events per thousand prescriptions dispensed, or six per million prescriptions dispensed. This is an extremely low number of reports, and further supports the acceptability of AB rated substitutable generic levothyroxine products.

In conclusion, Mylan supports the bioequivalence standards for levothyroxine established by the FDA. In response to recommendations put forth in previous citizen's petitions that were filed by name-brand manufacturers with regard to levothyroxine, the FDA added a requirement for baseline subtraction of T4, as you've also heard this morning, so that the endogenous levels of T4 in study subjects participating in levothyroxine could be subtracted from bioequivalence trials. Mylan accepted and agreed with the additional requirement, and considers the current FDA criteria to be acceptable for determining

that generic levothyroxine products are safe, effective, therapeutic equivalents to their name-brand counterparts. Thank you. I'd like to have Beth Brannan from Sandoz.

DR. ORLOFF: Beth Brannan from Sandoz to introduce your speakers.

MS. BRANNAN: Good morning. Getting close to `good afternoon' in fact. My name's Beth Brannan, and I'm the director of regulatory affairs at Sandoz. And I'd just like to thank FDA, the American Thyroid Association, the Endocrine Society, and the American Association of Clinical Endocrinologists for allowing Sandoz to have time to present today at this public meeting.

And I'm going to introduce our speakers, our panel of experts this morning. We have Dr. Robert Richards from Louisiana State University. He's going to give a provider's perspective. And Sally Schimelpfenig will give the generic market perspective. And Alfred Elvin will present our bioequivalence perspective. And Bruce Weintraub will provide comments on the clinical aspects.

We also had some additional people on our panel of experts that are not here presenting today. Dr. Les Bennett, who really doesn't need any

introduction, Dr. Sandy Bolton, and Dr. Tony Toft, a top endocrinologist from the U.K. So first up we have Dr. Robert Richards.

DR. RICHARDS: Thank you. It's a pleasure to be here. In the early part of my clinical training, my early experience, I initially wrote for generic thyroxine only. I did this for years. Then one day I started writing for brand name thyroxine. Why? Was it because my patients were not doing well? No. My patients were doing fine. I allowed a drug rep to overly influence me. Well, I continued this for a couple of years, and then I went full circle and resumed writing generic thyroxine. After a few years, I made an observation. My patients were doing fine. They were doing no better, they were doing no worse, whether they were on generic or on brand name thyroxine. My current view is that generic thyroxine is fine for patient care.

Today you will be hearing about TSH and free T4 being debated. Please remember that TSH varies inherently. It follows a diurnal rhythm where the peak is in the morning and the nadir is in the afternoon. Some investigators report that the difference between peak and nadir is about 50 percent. Despite this degree of variation during the day, I'm

not aware of many physicians instructing their patients to always get their TSH tested at a specific time of the morning.

Superimposed on this diurnal pattern is the pulsatility of TSH. We all know that pulsatility will greatly affect variation. Despite this, I am once again not aware that physicians are ordering serial TSH measurements in their patients during the course of the morning in order to minimize the influence of these pulses. Of course, the TSH assays themselves introduce variation, and there are other sources of variation in TSH. One problem is the patient who misses a dose. I know most of our patients try to be complaint, we try to believe our patients are compliant, but sometimes they will miss a pill. If they miss one pill during the course of a week, that is equivalent to a 14 percent reduction in their dose. Unfortunately, some of our patients miss more than one dose. They may go for a period of time without taking their pill, and then they realize. They come back to the clinic, and they'll start taking their thyroxine again. When they show up in clinic, their free T4 is usually recovered. Free T4 responds faster than TSH. TSH lags behind. Some cases, many weeks, sometimes six weeks or more before it reaches

its new level.

Intestinal absorption of thyroxine is affected by a number of agents as you've already heard. These include some prescription drugs, some over-the-counter formulations, and some dietary supplements. Despite our best efforts, we are never sure when or if our patients are mixing their thyroxine with one of these substances. Variability will always occur, whether the patient is on brand name or on generic.

We all care about patient welfare. Some will argue that good patient care requires brand name thyroxine only. A portion of this is explained by the Carr Study in 1988. I'd like to point out that that was 1988, long before the FDA has instituted this more rigorous verification of thyroxine doses.

Patients do well on generic. I successfully treat patients with routine hypothyroidism using generic thyroxine. Some of my patients have had thyroid cancer. I share the same concerns that many of the people in this room share, and that is that the TSH must be suppressed in these patients, but not overly suppressed. I can do that with generic thyroxine. Some of my patients are pregnant. We all know that the thyroxine needs of a

woman dramatically increase during pregnancy, not always in a predictable manner. Therefore, we follow these patients frequently, watch their labs, their clinical presentations, and adjust their doses as needed. I'd like to point out that even a woman who is maintained on the same brand name of thyroxine throughout her pregnancy would still need to be tested frequently because her dose will have to be modified.

Most of my patients are at Charity Hospital. Charity Hospital, and the other hospitals in the State of Louisiana are mandated -- at least the state hospitals -- are mandated to use generic thyroxine. It doesn't matter what we write for an inpatient. I have checked with some of my colleagues, and I have found that most of them prescribe generic thyroxine. They have not seen any change in patient outcomes, and they have not seen any need for more frequent follow-up. I have checked with some of my patients who are taking generic thyroxine. They all seem satisfied with it.

The American Thyroid Association, the Endocrine Society, and the American Association of Clinical Endocrinologists have published a position statement. Unfortunately, I feel that this position statement is a little biased against generic

thyroxine. I am a member of two of these organizations, and I can assure you that I have never received a draft copy of any position statement before publication, or given a chance to read and express my opinion for publication. I'm not sure if these position statements truly reflect all the views of the members.

In closing, most of my patients are indigent. Even though brand name thyroxine is relatively inexpensive compared to most drugs, it is still difficult to be afforded by patients with no job, no insurance, no financial support. This is not unique to New Orleans. Many people in this country are either uninsured or underinsured, unemployed or underemployed, poor or becoming poor. It is my feeling that routinely substituting generic thyroxine will help my patients. This will improve their compliance, and their expected outcomes. This saving is especially true for some of the my older patients, who are on multiple drugs. Generic substitution does not take control away from the physician. The physician can still write on the prescription pad `Dispense as written' or whatever phrase is needed in their state for those patients that he or she deems necessary.

In conclusion, inhibiting generic substitution will unnecessarily raise health care costs. Please do not change the current system. Please decide in favor of our less fortunate patients. They don't have the advocates that other groups enjoy. Thank you for your time.

MS. SCHIMELPFENIG: Hi, I'm Sally Schimelpfenig, in the marketing department at Sandoz. I'm the product director for levothyroxine, so one of the things I do frequently is to track where we are in this market, and post-approval the big question is what has changed. And what changed was we went from a market where there were two competitors to post-approval of the therapeutically equivalent products, we now have a market with five competitors.

As you can see, by increasing the level of competition in a market, you can bring savings to that market, big savings. And for a product that is as widely prescribed as levothyroxine, these savings are spread very evenly across the patient populations and the health care system. What we're looking at here is a savings of $145 million since launch. That's an estimated number of all generic product. And that estimated number is based on the substitution rate, currently at 25 percent, which is greatly suppressed

compared to other molecules that are genericized.

Another thing I'd like to be able to bring to your attention is the total units, annual units, of this product, estimated to be at about four billion units. I would also like to point out that the estimated total annual sales of this product are about $1 billion. That having been said, for every generic substitution that is made there is increased savings to the system, which greatly assists the system in being able to afford more innovative care for more critical states -- not more critical states. More innovative care for newer therapies, and be able to maintain patients safely on levothyroxine. Thank you.

DR. ELVIN: I'm Alfred Elvin, director of biopharmaceutics, Sandoz. Every current generically marketed levothyroxine product has been approved and rated by FDA as therapeutically equivalent, or AB rated, according to FDA's expert guidance. No authenticated data exists on any FDA-approved, therapeutically equivalent levothyroxine product demonstrating any difference in safety and efficacy profile between the approved AB rated drug and its reference-listed counterparts, and for that matter, any approved generic drug to date.

The three levothyroxine products approved

as AB rated are pharmaceutically equivalent to the reference-listed drug products. The three levothyroxine products approved as AB rated are bioequivalent to the reference-listed products.

Levothyroxine characteristics, summarizing what's been presented this morning. Levothyroxine is highly soluble. It's 100 percent dissolved in less than 30 minutes. The formulations, as indicated by Dr. Duffy, are made to current manufacturing specs, modern specs. They're reliable, direct compression.

Potency difference in Sandoz studies. The FDA requires that any product compared to a reference product in a bioequivalent study differ by less than 5 percent. In practice, our manufacturing matches Mylan's. Our differences in potency from lot to lot vary from 99 to 101 percent.

The FDA levothyroxine guidance accounts for endogenous plasma T4 variability through a baseline correction method which provides an appropriate statistical basis for FDA to define levothyroxine bioequivalence. Based on Sandoz submissions, the FDA determined that Sandoz levothyroxine is pharmaceutically equivalent to the reference-listed products, bioequivalent, and therefore, therapeutically equivalent, AB rated, to

both reference-listed products. Thank you.

DR. WEINTRAUB: Thank you very much. I'm Bruce Weintraub, and I've been in both worlds. I've been in the clinical academic world, and now I'm in the biotech world. And I think I have a unique perspective on both sides of the issue. I've been in TSH research for most of my life. I've worked with my distinguished colleague Chip Ridgway many years ago on the development of the sensitive assays that permit the kind of monitoring we're talking about. I've worked on all aspects of TSH physiology. I was the inventor of recombinant TSH, which is used for other purposes in working with my colleagues. In the course of that, I worked with the endocrine metabolic team at FDA, and I got an appreciation of FDA standards of pharmacokinetics and bioequivalence that clinicians may not always appreciate. And similarly, in my current biotech company, I'm always dealing with these issues. So I really think I have a balanced view of it.

And I want to say that being in both worlds, having the balanced view, I come down heavily on the side of the FDA, that the FDA current NDA standards are the appropriate ones. Because although TSH, which is very dear to my heart, is usually a

sensitive measure of thyroid function, as you've heard it's an indirect measure and has limitations. You've heard from my colleague some of the limitations. I'll add to it. There are other factors, non-thyroidal illness, central pituitary or hypothalamic hypothyroidism, psychotropic drugs, heterophilic antibodies, many things influence this. Clinicians are used to dealing with the limits of TSH, and do a fine job of managing hypothyroidism associated with these conditions using T4, free T4, and clinical indices.

TSH is an invalid drug bioequivalence measure as a result of intra-patient variations. We haven't heard enough about the variations that occur even in the same patient on a branded product. Enormous variations, mostly due to compliance, weight, all these things. It's not as stable. The variation that might occur from a switch, if there is any at all, would be dwarfed by these intra-patient variations. And it is therefore not an appropriate indirect measure.

Moreover, T4, or free T4, is the direct and accurately, and easily measured analyte. And it is the most meaningful clinical measure of drug absorption and bioequivalence using conventional FDA

standards. The FDA has an enormous history, as you have heard, of doing bioequivalence. When the analyte is measurable, as it is so easily here, they always choose to use the direct analyte because of problems of indirect measurements. This has stood the test of time over decades and many drugs. There is no reason to change these time-proven criteria for L-thyroxine.

This is an old therapy. There are no IP issues here. The branded companies played no role in the development. There's no protection of IP that's relevant at all. As you heard, it's soluble, easy to measure, easy to manufacture, and these new NDA standards are really going to, I think, protect the public.

Now, I want to emphasize in closing two points. The current standards of care call for routine lab value monitoring of TSH, with or without T4, free T4, at least once or twice yearly. And that's taking into account, again, variability even of patient on the same level. So such monitoring, if adopted, and I strongly recommend it, not unique for the generics, or not switching, but just in general, because of intrinsic variabilities of patients' TSH. I think it provides adequate safeguards to prevent chronic, and I emphasize chronic, over- or under-

treatment, and greatly mitigates any threat of long-term health risk from exogenously induced hypothyroidism and hyperthyroidism.

And then finally, consensus views, and I stress consensus because there's a lot of debate about this entity, but consensus views of thyroidologist relating to the clinical significance, clinical and metabolic significance of so-called sub-clinical hypo- or hyperthyroidism, which is a decreased or increased TSH with normal T4 or free T4, are associated with TSH values well above or below the normal range for periods of many years, or even decades. And I'll get into more description of that. Such extreme TSH values for such long periods would not be encountered in patients switched to generics, and receiving recommended monitoring. Thus there is no convincing evidence for claims -- and I think they're dogmatic claims, they're not supported by the evidence -- of such an ultra-narrow therapeutic range for thyroxine therapy. And in any case, even if there were, such claims would have to take into account the duration of such therapy, and how difficult it is to prove metabolic impact of these changes when they're not studied in large numbers of patients over years or decades.

And I want to just close with an anecdote because I in my academic world, I had a lot of the prejudices of the clinicians, and I thought that it was an ultra-narrow range. But then I did a study with Jean-Jacques Staub from Switzerland on the metabolic, and I emphasize the metabolic impact. It's not just the TSH. The Carr Study quoted in a small number of patients did not look at the metabolic impact. But we looked at a very large number of patients with so-called sub-clinical hypothyroidism over many, many years and decades. And we could only demonstrate a metabolic impact, and a clinical impact, with TSH over 12. You notice on the slide from the Abbott gentleman, he was talking about increased risk of hypothyroidism, clinical consequences, when it was above 4. But the data don't support that there are clinical impact until you get quite high values for very long periods of time. So I then saw that I had prejudice and bias that was not supported by the data; that if you really look at the metabolic data, that it has to be extreme.

And Dr. Ladenson pointed out to me that we did not study the opposite, and I don't have the same experience, but from looking at the literature, I would believe it would be the same, that these small

generic substitutions would produce, even if Dr. Sherman designed a beautiful, perfect study, and because of the sensitivity TSH got a small and even significant difference, I would believe you could not show any metabolic impact. And same in treatment of hyperthyroidism. Most of the statements about the need for titrating the TSH at a certain level for hyperthyroidism, I'm balancing them, they're pure prejudice. They're not supported by prospective trials looking at metabolic impact beyond TSH.

So I go back to the bottom line. The proof is in the pudding. These generics have been out now for quite a long time. You've heard from distinguished clinicians with large numbers, we're talking here over one billion -- this is the Sandoz product -- one billion products released, 43 million prescriptions, very small number of adverse events, non-serious events, events that in placebo-controlled trials would be an equivalent number of non-serious events. And distinguished clinicians in states like Louisiana who have no control over substitutions, they honestly cannot tell the difference, not only clinically, but in the total and free thyroid hormone levels and TSH levels. So despite dogma that I used to share with my clinical colleagues, when I really

look objectively from my new biotech perspective and working with the FDA, I come down heavily on the side of the FDA and generics, and feel these are appropriate standards, and no patient will be put at risk by substitution with generics. Thank you.

DR. ORLOFF: Okay, thank you very much. It is now five minutes of 12:00, and we are going to break for lunch. And I'd like people to return here by 12:50 so that we can have another half an hour of public comment period, and it's hoped some panel discussion. So the morning session is adjourned. We'll see you at 12:50.

(Whereupon, the foregoing matter went off the record at 11:56 a.m. and went back on the record at 12:57 p.m.).

DR. ORLOFF: Why don't we get started with the public comment period. We have approximately 30 minutes. Because a number of people have asked to speak, I'm going to need to limit everyone to three minutes during this comment period. There will be a yellow light in front of you on the clock with one minute to go. The first speaker is Dr. Garber from the American Association of Clinical Endocrinologists. You can come up front, it's fine. The next speaker is Dr. Alan Farwell from the ATA. So I'm going to

have the people in the on-deck box. Go ahead, Dr. Garber.

DR. GARBER: Three seconds into my time limit. I'd like to thank you, as everybody else seems to be thanking you, for appearing here today. I'm Jeffrey Garber. I'm a clinical endocrinologist. I live and work in the Boston, Massachusetts area, and I'm currently the secretary of AACE, the American Association of Clinical Endocrinologists who I'm representing today. AACE has over 5,000 members. Virtually all of our members are practicing clinical endocrinologists. My own practice over years has enabled me, or given me the opportunity to take care of and continue to care for literally thousands of people with thyroid disorders.

What I'd like to address is give you really two concrete examples of how this issue can affect patient safety. The first is if we extrapolate from the Carr data, and what I've heard repeatedly today, and seen in print, that a Sandoz preparation may in fact be 12.5 percent more than Synthroid, the issue not only is a 12.5 percent difference in dose, which is often 12 or 13 micrograms or more, it's whether when you switch somebody from one preparation, because you've increased their dose by 12 or 13

micrograms, and then you have an additional variability of an additional 12 or 13 percent, you're basically dealing with 25 microgram differences. And if one actually looked through the Carr data, it's not only as it's represented. It actually under-calls a very important point, which is there wasn't a single patient in that study who you couldn't change their range of control by switching them to 25, if you just went through every part of the spectrum. So you take a frail elderly person who is prone to atrial fibrillation, and as opposed to bone disease and the like, cardiac events can be fairly acute, and often fatal, and we don't really necessarily monitor people in any kind of routine fashion with that kind of frequency that we could know that. And that's one major concern, vulnerable elderly cardiac patient. And even someone who's not that elderly.

The second one is actually -- hits a little closer to home. Sub-clinical hyperthyroidism and hypothyroidism is by definition impossible to clinically diagnose. What happens is we see somebody and we say `You're perfectly fine, we just checked your levels, we've fulfilled every kind of monitoring criteria imaginable,' and they call us up a few weeks later and say they feel lousy, or they have depressive

symptoms or palpitations. Well, that compels us to re-check them, but more than just the money, and the cost, and the inconvenience of doing that, the thing I fear the most, it actually leads to potential for delay and misdiagnosis. These people may be having a primary cardiac problem that has nothing to do with their thyroid, or they may be having depression, and we just don't tend to them soon enough. So this is another smokescreen that a busy clinical practice has to contend with, and I think we should do what we can to eliminate these kinds of manageable variables. Thank you.

DR. ORLOFF: Dr. Farwell from the American Thyroid Association. The next speaker will be Dr. Lawrence Wood.

DR. FARWELL: Thank you very much. My name is Alan Farwell. I'm a clinical endocrinologist and associate professor of medicine, and director of the endocrine clinic at the University of Massachusetts Medical School, and council member of the American Thyroid Association, the organization I am representing here today.

The American Thyroid Association, also known as the ATA, is a society of physicians and research scientists founded in 1923, and is a leading

professional organization dedicated to the thyroid. Our mission includes promotion of thyroid research, improving diagnosis and treatment of thyroid diseases, and education of professionals and patients about thyroid disorders. Our website, thyroid.org, is a leading provider of clinical thyroid disease information on the internet, and receives over 1.5 million visits per year, mostly from thyroid patients seeking educational information about hypothyroidism, the disorder that is treated with levothyroxine.

I want to emphasize that the ATA, just like AACE and the Endocrine Society, is not against lower costs of medications, it's not against lower -- decreased access to care, and not against any specific generic or branded thyroxine preparation. We are for precise dosing without significant variation for our patients. In 2002, we organized the ATA Alliance for Thyroid Patient Education, which I chair, and which consists of the major patient education and advocacy organizations in the United States, including the Thyroid Foundation of America, the Thyroid Cancer Survivors Association, otherwise known as ThyCa, Light of Life Foundation, and the National Graves Disease Foundation. The members of these organizations are thyroid patients as their main membership, and they

are the constituency which we serve as physicians. You'll be hearing from representatives of two of these organizations later on today, Cherry Wunderlich from ThyCa and Larry Wood from the Thyroid Foundation of America. There is a strong concern among these patient groups that the ability of the physicians to prescribe and monitor their thyroxine therapy has been compromised by the FDA decision in last June of 2004.

Three major issues have become apparent since last June. Number one, many patients have been switched to generic levothyroxine products, did not know they had been switched, and that will be discussed a little bit later on today. In many cases, managed care organizations have substituted their generic products for lower tier coverage and pushed the brand products to their highest tier. So there is no cost savings to a patient going on the generic products, but there is a significant increased cost for patients who wish to stay on a branded preparation. Indeed, there are some insurance companies that will only provide the generic. And third, most patients that have been switched to generic levothyroxine products, in contrast to the FDA's goals, have been required to get a dose change. In my own practice, a review of the last 21 patients

that were consecutively seen by me that were switched from a branded preparation, 18 required a dose change. In short, the approval of the current generic levothyroxine products has not provided any advantage to the patients being on these medications. On the contrary, they have led to more unintended symptoms, more doctor visits, increased non-pharmaceutical health care costs, and significant disruption in patient's health and wellbeing. Thank you very much.

DR. ORLOFF: Dr. Wood, and the next speaker will be Dr. Rosalind Brown.

DR. WOOD: I'm Larry Wood. I practice in the thyroid division at the Mass General Hospital in internal medicine. With the help of several patients and colleagues in the thyroid unit, 20 years ago we created the Thyroid Foundation of America because we thought patients needed to be educated better and supported to understand what was going on when they got a thyroid problem. One of the things we have done for the last 15 years is we've had a patient, or a woman, an educated thyroid specialist talking to patients on the phone and answering any questions they have. Everything we do is free.

About six months ago, Ellen began to get increasing numbers of calls from patients who were

concerned about having to change their thyroid medication. We responded, and then we decided we ought to be a little more scientific, so we started a survey on our website. I just wanted to summarize the two most significant aspects of that survey so far. Of 159 patients who were changed, 50 percent, or 76 were changed not by the doctor, not by the nurse, but either the pharmacist or because of insurance company regulations. Secondly, our patients had been educated that they should -- if they changed, they needed a follow-up TSH test to be sure their dose was correct. Of 159 patients, 111 had abnormal TSH tests, or 70 percent when they were re-checked, 25 percent were hyperthyroid, and the rest hypothyroid. So I speak on their behalf asking you to listen to what patients are saying. They want to be part of the picture, and they're scared to death that they're losing control. Thank you.

DR. ORLOFF: Thank you. Dr. Brown? Dr. Brown? And the next speaker will be Cherry Wunderlich from the Thyroid Cancer Survivors Association.

DR. BROWN: My name is Dr. Rosalind Brown. I'm an associate professor of pediatrics at Harvard Medical School, and director of clinical trials research in the endocrine division at Children's

Hospital in Boston. I have devoted my entire professional career to the study and care of children with a variety of thyroid diseases, and I'm here today to represent the Lawson Wilkins Pediatric Endocrine Society, which is an organization of approximately 800 pediatric endocrinologists who are dedicated to the care and study of infants and children with hormonal disorders.

Today we have heard a lot about various methods of determining bioequivalence. My purpose is to persuade you to think about a particularly vulnerable population that we have not yet mentioned, and to convince you why we must not be satisfied with anything but the most sensitive markers of bioequivalence. Approximately 1 in every 3,000 infants born each year in this country and elsewhere suffers from thyroid insufficiency, a condition known as congenital hypothyroidism. As recently as 30 years ago, congenital hypothyroidism was the commonest treatable cause of mental retardation in this country.

Due to the realization that the IQ of affected infants was related to how early thyroid hormone replacement was started, newborn screening programs for the detection of congenital hypothyroidism have now been detected not only in

North America, but throughout the world. These programs have been dramatically successful in eradicating the mental retardation caused by this disease. However, it has become abundantly clear that the cognitive outcome of affected infants depends exquisitely on the dose of thyroid hormone replacement used. A difference in starting dose between 8 micrograms per kilogram, approximately 25 micrograms for the average infant, and 10 micrograms per kilogram, approximately 37.5 micrograms, has been repeatedly associated with a significant difference in IQ. What this means in practical terms is that substitution of a different formulation of thyroid hormone that is not precisely bioequivalent can have a devastating effect on the infant's outcome if the physician is not aware that this has occurred, and thyroid hormone has not been re-titrated. Furthermore, because of the critical window of thyroid hormone dependent brain development, if for example a physician only learns that the thyroid formulation has been switched two months later, the consequence to the infant is irreversible. This is quite different from the subtle adverse effects that you have been hearing about which take years to manifest. It is estimated that something like three to four IQ points are lost

for every one to two microgram difference in T4.

In summary, babies with congenital hypothyroidism are an example of the smallest and most vulnerable patient population who demonstrate the narrow therapeutic range that is necessary for optimal thyroid hormone therapy. The present methodology employed by the FDA in determining bioequivalence, although a significant improvement from methods in the past, remains insufficiently sensitive and precise, and as a consequence can have serious, irreversible consequences to our infants and children. The Lawson Wilkins Pediatric Endocrine Society feels strongly that evaluation of bioequivalence should be changed to one that considers measured levels of TSH, which is the universally accepted standard of care in thyroid hormone therapy. Thank you.

DR. ORLOFF: Cherry Wunderlich? And Peter Lurie is the next speaker.

MS. WUNDERLICH: Thank you for this meeting. I'm from ThyCa, Thyroid Cancer Survivors Association. I'm Cherry Wunderlich, ThyCa board member. I'm giving this statement for our board chair, Gary Bloom. We're thyroid cancer survivors and ThyCa volunteers. As thyroid cancer patients, we have serious concerns about the matters being discussed

today. ThyCa is a national nonprofit organization advised by nationally recognized thyroid cancer specialists. ThyCa provides free education and support for patients, families, and the public. Our services include support groups, publications, workshops, and conferences. We have 5,000 to 10,000 participants in our support groups alone. Our website receives more than 200,000 hits each month.

The need for patient support has grown rapidly because thyroid cancer is one of the few cancers that is increasing in incidence. We urge you to use the guidance of the leading endocrinologists on the crucial issues related to levothyroxine sodium bioequivalence. These endocrinologists are experts on thyroid issues and thyroid patient care. We patients benefit every day from their knowledge and expertise. We greatly appreciate their dedication to patient wellbeing. Like other thyroid patients, we need to be sure that our blood levels of thyroid-stimulating hormone, TSH, stay at the target level needed for our individual circumstances. A precise TSH level helps prevent growth or recurrence of the most common types of thyroid cancer. Dose changes prescribed by our physicians are small, even tiny, usually less than 10 percent. For these reasons, our website's Know Your

Pills page explains key points about levothyroxine, and the advice of our specialists, to avoid changing brands without being re-tested for TSH level.

In addition, regarding bioequivalence studies needed, with over 300,000 thyroid cancer survivors, all of whom are dependent upon thyroid hormone for their survival because they have no thyroid gland remaining, we are confident that more than enough thyroid cancer survivors would volunteer to participate in needed bioequivalence studies. We strongly support the analysis and recommendations of the leading endocrinologists in the American Thyroid Association, American Association of Clinical Endocrinologists, and the Endocrine Society. As patients, we ask you to support their recommendations. Thank you again for your time and consideration.

DR. ORLOFF: Thank you. Peter Lurie? And then Sally Schimelpfenig is welcome to come up as well for the last three minutes.

DR. LURIE: Good afternoon. I'm Dr. Peter Lurie, deputy director of Public Citizens Health Research Group. Coming to this hearing today is a little bit like attending a showing of the movie Groundhog Day. This hearing is simply the latest round in a decades-long debate in which discredited

scientific arguments, be it the Carr Study which we've seen a million times before, the Blakely Study we've seen a million times before, are repeated, added together with uncorroborated clinical anecdotes. And the only real new wrinkle here is that instead of the arguments coming only directly from the company, they come instead from the three major endocrine societies, all of which, if you look at their websites, take significant funds from Abbott. I also wish that some of the previous speakers had disclosed their conflicts of interest. I for myself, Public Citizen, we take no money from government or industry.

So, here is a meeting completely set up that would otherwise not happen were it not for the force of the companies acting either directly or indirectly, and they have been successful. They have hung on in the case of Synthroid to 82 percent of the market, even though Unithroid sells for half the price. In comments that I'll submit to the record, we estimate that this costs the American consumer over $200 million every year in the absence of any clinical benefit. Part of the problem here is that there are now a plethora of these formulations on the market. There are eight of them at least listed in the Orange Book, which means there are 28 combinations of drugs

that might be tested in pairs for bioequivalence. Only seven of these have been done. And so Drug A is similar to B but not to C. Everybody's very confused by this. I think an important role for the FDA is an educational one, to explain to the pharmacists what has legitimately been shown to be substitutable. I also think that some of the holes in that matrix with the 28 combinations could be plugged if the Agency for Healthcare Research and Quality were to use its Centers for Education and Research on Therapeutics, or CERTs, to actually conduct some of the bioequivalence studies and get rid of some of the uncertainty.

Part of what Abbott is trying to do is to exploit, again, the TSH. And as it well knows, TSH levels are subject to a number of influences, many of which have been outlined today. We also know that TSH behaves in a distinctly non-linear fashion. The changes at the lower end of the spectrum are very different than a similar change at the upper end of the spectrum. It's exactly that source of noise that the company is trying to exploit, knowing full well that it would result in a requirement for massive sample sizes in any effort to prove bioequivalence. In fact, Dr. Conner of the FDA, when speaking at the March 2003 advisory committee meeting -- that was the

previous Groundhog Day -- he said, "In fact, I would 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 in the TSH."

As it happens, there's a far more fundamental question, which is whether or not TSH is a reliable predictor of clinical outcome at all. Dr. Anthony Toft, who I gather was supposed to be here, stated in a recent editorial, quote, "There is simply no evidence, other than anecdotal, that an increase or decrease in thyroid tablet content of up to 12 percent will induce sub-clinical or overt hyper- or hypothyroidism." And as has not so far been mentioned, there is an important article in the Journal of the American Medical Association of the last year or so in which these same three societies requisitioned a meta-analysis of all the data on sub-clinical hypothyroidism and found the following results. The review found that the available data were, quote, "insufficient to show a benefit upon lipid levels, cardiac dysfunction, systemic

hypothyroid symptoms, or neuropsychiatric symptoms from treating patients with TSH's of either 4.5 to 10, or even over 10 million international units per liter." Furthermore, the review found no evidence that treatment of either of these TSH levels had an impact upon adverse cardiac endpoints. TSH is an important clinical tool. It is not a useful bioequivalence tool.

Finally, the companies actually are asking the FDA to break the law with respect to the involvement of TSH in the determination of bioequivalence. As we've seen before, there is a hierarchy of different studies. But what was not mentioned by the FDA speaker is that it's made clear that you're supposed to use the top of that hierarchy, and not the third of the hierarchy, which is where TSH would fall. The regulations permit this less desirable third approach, quote, and I'm quoting from the regulations, "only when appropriate methods are not available for measurement of a concentration of the moiety, and when appropriate it's active metabolites." Clearly that's possible here, so Abbott is literally asking the FDA to break or rewrite existing regulations, regulations that has served us well.

I guess I'll close with a quote from Groundhog Day. Phil, that's the character played by Bill Murray, who says, "Well, what you do if you were stuck in one place, and every day was exactly the same, and nothing that you did mattered?" Well, that about sums it up for me. Thank you.

DR. ORLOFF: Thank you. Are there any other? Dr. Schimelpfenig?

MS. SCHIMELPFENIG: I'm going to waive.

DR. ORLOFF: You're going to waive? Okay. I'm going to turn it over to Dr. Ladenson. And I hope in the next public comment period we'll get some time for actual questions from the audience, and questions from the panel so that we can engage in discussion. Dr. Ladenson?

DR. LADENSON: Thanks, David. The next speaker is E. Chester Ridgway, who's Director of Endocrinology at the University of Colorado Health Sciences Center. Dr. Ridgway is going to talk about the rationale for TSH as a marker of thyroid hormone tissue effects.

DR. RIDGWAY: Thank you for the opportunity to give this talk. I'm here to talk about TSH, and try to defend the TSH as a useful and absolutely mandatory monitor for future bioequivalence

studies. I'm going to make four points. We'll start with the first. TSH is the most sensitive measure if thyroid hormone action. I believe that that is clinical wisdom as well as over a thousand studies to show that.

TSH is a pituitary glycoprotein hormone. It controls thyroid gland growth, function. TSH production and secretion are very sensitive to circulating thyroid hormones, and as mentioned earlier, the TSH secretion is pulsatile and circadian. Mean pulse frequency is 7 to 13 pulses per day, and amplitude, meaning the height of these pulses averaged over a 24-hour period is 2.5, but in the daytime it is 1.5 to 2, and the mean nighttime is a little bit higher. This is a typical pulsation of a normal control. You can see the pulses asterisked. I think this person has 11 or 12 pulses in the 24-hour period. You can see that they all lie within the normal range for the TSH assay. Most importantly, you can see that during the daytime hours, the pulses are quite low in amplitude. They span a difference of approximately 0.9 to 1 microunit per ml. We do not get huge high pulses in the morning. The times alluded to earlier today were a little bit off. The peak starts at 11:00 p.m. and ends usually at 4:00 to 5:00 p.m. in the

afternoon. There are no peaks in the daytime hours when we actually do clinical practice.

Here is another patient. This one is on levothyroxine showing you exactly the same kind of a pattern, all within the normal range, peak in the evening. All of them reside with this very small amplitude change of 1 to 1.5 microunits per ml. This is a very, very steady pattern, and these do not vary all over the map as implied earlier.

This is a study of Andersen that actually showed basal levels of TSH over a year's time, 15 normal euthyroid controls. And each of these dots signifies one month TSH value. And you can see that they're ordered from lowest to highest. You can see that there is low variance down here in the low levels, a little bit higher variance up in the high levels. Again, note the scale that these do not vary over 1 to 1.5 to 2 microunits per ml. Now, are each one of these pulses, like this one right here, is that a pulse? Or is that because of some seasonal variation? The study hasn't been done. We haven't done 24-hour curves, 12 times the normal controls. Or are all of these pulses? This is easily testable. Would all of these even out into the same pulse pattern if you actually did the study? We need to do

that before we make claims about irregularities and inaccuracy of TSH measurement.

In this particular population, this reference group defined a new normal range for this group. And you can see that its mean is lower. This is important because this is what this looks like as far as the reference population is concerned with any normal reference population of TSH. In this, the Denmark group had this new reference range for its 15 normal people. One individual of those 15 would have a normal pattern that would consume about 50 percent of the reference population. The next patient would have a little bit different one, and every single one of the rest of the patients would have something different. And what we need to find out is whether over a 24-hour period these same kind of differences in areas under the curve for TSH are the same. It's a study that should be done before we make claims.

As you all know, there is a very sensitive inverse relation between the log of TSH and free T4 or T4. This is the paper of Spencer that has actually catalogued this, very log linear. And I think the important point here is that for a twofold change in free T4, you get a hundredfold change in TSH, or a 1 to fifty-fold difference. This is extremely important

as far as the sensitivity of TSH for monitoring therapy.

Second point. Normal thyroid hormone levels are not accurate measures of normal thyroid hormone action. So what do we mean by that? This is a figure from Dr. Wartofsky, in a review. One that is well taught in every single medical school. As you progress from euthyroid to mild thyroid failure, the hypothyroidism, the earliest sign of that failure is the TSH, which jumps out right at the beginning of mild thyroid failure. As a reminder, thyroid hormone levels do not change during that period of mild thyroid failure, and they all stay within the normal range. And this is the area that is so important. How many of our patients with thyroid gland failure actually fit into this group? That comes from -- one source of this study is the Colorado study, NHANES is the second source of this. They all show the same thing. The prevalence of a high TSH in this study being over 5.1 is 9.5 percent of the Colorado population. This is the largest study that's ever been done to study this. Those are the low TSH's, 2.2 percent or about four- or five-fold, less prevalent.

Now, how many of these actually have normal thyroid hormone levels? Ninety-five percent of

them have normal thyroid hormone levels. Ninety-four percent of patients with low TSH have normal thyroid hormone levels. This is a big population. It's an important population, and it's the one that we're trying to do well with as far as our patients are concerned today.

Third, past bioequivalence studies using T4 have made mistakes. Obviously, these studies were done before the current evaluations of bioavailability, the current drug, but it illustrates a very important issue. These mistakes would have been predictive that TSH has been included in the formula. And I'll show you that. Blood T4 levels are not the active ingredient, and they are not being measured at the site of action. Two very important criteria for FDA.

So this is the famous Dong study, presented in JAMA, 1997. And these are the bioequivalence. Notice here that the bars are a little bit narrower than what we're talking about today. The area under the curve, T4, two of the branded products that are being discussed today, and two generics which are not the two generics talked about today that have been represented. And as you can see by their uncorrected bioequivalence standard

FDA rating, these individuals were all rated as bioequivalent. And you understand the storm that that led.

Well, baseline correction, after the Blakesley Study occurred, this is what the bioequivalent -- none of them were bioequivalent. Every one of them were off base. Now, the reason for showing you this is not to show you how important correction it is. It's to show you that TSH would have done the same thing for you. And that's shown in this slide. If you actually measure the area under the curve for TSH's in these various combinations and comparisons of the drug, none of them would have been bioequivalent. All of them would have been off. And these are uncorrected TSH values. If you actually correct TSH values, it gets worse, the story gets even more convoluted, and more difficult to understand. So TSH, if they had been used as an area under the curve in this study would have predicted non-equivalence.

I want to show you a few specific examples of this. Just show you the enormity of what this is. So what I'm going to show you here now are T4 levels, and TSH levels over the 24-hour periods of the four drugs combined in a given patient. So here's the first patient. One individual, four different drugs,

LT4 levels, over the 24-hour period. To me these look pretty good. Look like they're right on target. And in fact, the TSH's look pretty good too. I don't think there's an endocrinologist in the room here that would quibble about this. These would be pretty good. They would have been thought to be bioequivalent. Now this is one patient in that study.

Here's the next patient. Again, T4's look terrific. TSH's, really bad. One TSH, note scale, starts in the twenties. Only the green line is normal for the TSH, where it should have been. The other two, completely suppressed. Three of the four would have induced a dose change in any clinical practice in the country.

The next one, another example. Again judge bioequivalence by T4. Look at this green line, though. Remember the rule, the tenfold, the fifty-fold, the hundredfold increase. Look what happens when you do the TSH. Not one of them in boundaries. One way above 20, all the rest completely suppressed. Every one of these would have required a dose change.

Now, am I being unfair by showing you three specific patients that tend to show the point? And I don't think so. Here is a summarization of that data. So Period 1, Period 2, Period 3, Period 4. If

you look at the mean TSH's, these are just the mean basal TSH's, not significant for any of these, when you look at just the means, comparing them in the group analysis. But if you actually break it down to who is high, who is low, and what are the combinations of abnormal TSH's for each period, 38 percent, 43 percent, 52 percent, 52 percent. Half the time the TSH's were not in range when a switch was made. And so I do not think that this is an exaggerating claim.

I would actually very much like to do the study that Peter described a moment ago. I think it would be very revealing to see whether same brand, done over a consecutive period of time, would give you this kind of data, or actually would give you more consistent data. That's a study that hasn't been done. They ought to include TSH's in that study when they do it, so that they can actually have the data. We wouldn't be guessing or making judgments without data.

Now, why is this? The problem is that we have a very complicated metabolism of T4. And it's different for different individuals, and it's different for different sites in the body. Obviously, this is the molecule thyroxine. There's an activation packed away, and two extremely important novel

molecules that we're just beginning to understand, the deiodinases that activate this pathway. There's also an inactivation pathway, and yet a third deiodinase, which is important for that particular process, to inactivate the hormone. And obviously the switch can occur when you actually go to diiodothyronine and the metabolic inactive product.

Now, what about these things, and why is this such an important thing to emphasize? Because I believe some of the variability that we see patient to patient is because of this. This is a schematic of thyroid hormone action. We all know that thyroxine hits the bloodstream, gets converted either in the plasma to T3, and if the cell gets converted ends up in the nuclei of cell, where it regulates gene transcription, either up or down, metabolic products in the form of proteins, or metabolic action occurs after that occurs. So, one important point is that D1 is largely an extracellular protein doing this in the extracellular space, whereas D2 is largely an intracellular protein actually doing this inside cells. Different tissues have different amounts of these deiodinases, particularly D2. So, the idea of measuring T4 as the only measure of bioequivalence is at least flawed in the first degree because it is not

the active ingredient. T3 is the active ingredient, and it's the thing that accounts for the thyroid hormone action. As I've been reminded many times, there are no intracellular events that we know that can be described by T4 at the level of the nucleus. Only T3. T4 is not the active compound. Likewise, the site of action is in the nucleus. The site of action is not T4 in the plasma. So two of the big rules, active ingredient at the site of action are both flawed when you deal with thyroid hormone, an endogenous hormone.

Finally, the toxicities of excessive or deficient thyroid hormone levels are now defined by TSH levels, not by thyroid hormone levels, not by thyroxine. To illustrate this in the past, thyroxine toxicity was defined by the clinical presentation, and secondarily by T4 and TSH levels. Let me give you an example of that. This slide of Graves Disease, the big toxicity not only -- but thyroids and a 50 percent chance of death. And here you'd have very high T4 levels, a suppressed TSH level, and that would be your definition. On the other side of the coin is in hypothyroidism, overt hypothyroidism, very low T4's, high TSH's, toxicity here is myxedema coma, in addition to the symptoms, and again, 50 percent

mortality here. This is what we used to do in the past.

Today, currently, thyroxine toxicity is defined only by the TSH level. And to give you that example, here is the example of sub-clinical hyperthyroidism, where the TSH goes outside the normal range, gets suppressed, whereas T4, T3 stay within the normal range. What are the toxicities here? Bone loss, fractures, myocardial dysfunction, cardiac arrhythmias, and death. I don't think Tony Toft is correct that there's been no toxicities associated with sub-clinical hyperthyroidism. Likewise, in the case of sub-clinical hypo, again, T4's stay within the normal range, TSH's go outside the normal range, and the toxicities here, decreased fetal IQ, increased lipids, abnormal vascular function, atherosclerosis, death, thyroid cancer recurrence and death. All of these have been alluded to.

I want to give you a few examples of these, and more examples will be given to you in a few moments. Let's take osteoporosis and fractures. This is a big prospective study from San Francisco, 686 from a cohort of over 9,000 women, elderly women, all adjusted by multifactorial analysis for previous hyperthyroidism, age, self-rated health, estrogen use,

and thyroid hormone use. TSH was low. Eighty-six percent of these people were on thyroid hormone. And what are the data? Here are the adjusted relative risk ratios for hip fracture and for spine fracture. The TSH is low. You get this dramatic increase, highly significant increase in fracture rate. This is not just osteoporosis now. This is fracture rate. Likewise, if the TSH is even minor decrease, a 0.1 to 0.4, it turns out that spine fracture is also significant also in this study.

Sub-clinical hyperthyroidism and atrial fibrillation. You've seen this study earlier today broken into the categories of TSH. Again, the toxicity of T4 defined by the TSH level. Same data, normal people set at 1. If you have a low TSH below 0.1, second generation assay, you get this 3.1-fold increase. Turns out that even the minor low levels hits right on our usual standard for significance at 0.05. And quantitating that into something real for clinical practice, it means that 28 percent of these people will get atrial fibrillation over a 10-year period of time. I submit to you that's a pretty heavy dose.

And does it have a clinical effect? Here's the Parle study from Great Britain that

actually measured TSH's, and then looked at survival and death. And the most important part of this curve is this curve, which actually defines death from a suppressed TSH level of less than 0.5. And I would like to say and remind Bruce on this, this is not decades. It actually becomes significant at the 2-year time point. It's significant at the 5-year time point. It doesn't take 10 years for this to occur. This occurs quickly, and can be quite devastating.

Minimally elevated TSH and lipids. This is the most recent study. The old Staub study is not the most recent study. This is the most recent study of 45 sub-clinical hypo patients. The TSH's here were not greater than 12, mean TSH's were 6.3. Most of them were in the 5 to 10 range compared to controls. This was part of a blinded RCT. I won't give you the RCT part of this, which was significant. To remind you that controls were definitely different as far as total cholesterol and LDL cholesterol. These changes were significant. As more recent studies come on, this has been the rule of thumb. Just a reminder about the Colorado study, 5 to 10 was also significant at 0.003.

Does it mean anything? To the heart, sure. Carotid artery intimal thickness, here it is as

a marker. Again, significantly different in sub-clinical hypothyroidism. Rotterdam study as far as long-term follow-up. This is only a cross-sectional study, 10.8 percent at a high TSH. MI, aortic calcifications were the toxicities. Set up 1 for the euthyroid group. Here's with an elevated TSH, and here's with elevated TSH plus antibodies. All of these significantly different.

And finally, the minimally elevated TSH and cardiovascular disease and mortality. This is the Japanese study, just out in JCEM, 2,500 survivors of the atomic bomb, 10 percent had an elevated TSH, 96 percent were within 5 to 10. Overall cross-sectionally, odds ratio, 2.7 for coronary artery disease significant. Men, 4.5 percent, odds ratio significant. Women not. All independent of other cardiovascular risk factors. And here is what the men looked like in follow-up over this 10-year period of time. Women not yet significant. Men becoming significant between the second and third year. It doesn't take decades to do this.

Conclusions. TSH is the most sensitive measure of thyroid hormone action. T4 levels are not sensitive to pharmacodynamic measures of LT4. TSH is the most sensitive pharmacodynamic measure of LT4, and

our plea is that TSH should be used in combination with total T4 for future analysis of LT4 bioequivalence. You will finally get a good complete picture of what these different agents are doing. Thank you.

DR. LADENSON: Thank you, Dr. Ridgway. The next speaker is Dr. Steven Sherman of M.D. Anderson Cancer Center, and the University of Texas in Houston. Dr. Sherman is going to talk about levothyroxine or TSH for determination of bioequivalence study design considerations.

DR. SHERMAN: Thank you for the opportunity to speak. I come from an institution where we take care of about 2,000 patients with thyroid cancer each year, and I would love to share with the you the story of a patient of mine with metastatic disease that progressed after a formulation switch, but of course that would just be an anecdote and of less import today.

What I will be talking about are some of the issues, both theoretical and have been demonstrated in published studies, about limitations of bioequivalence testing, and how one might design perhaps what I think would be a better form of bioequivalence study. The heart of it comes down to

switch-ability. And the reason that FDA cites for their approach to bioequivalence testing is to assess the relative bioavailability between test and reference product, permitting therapeutic equivalence. And as cited in a recent publication of which two the FDA panel members were coauthors, these measures of systemic exposure, including AUC and Cmax are assumed to relate to clinical benefit endpoints.

Now, as a clinician, my perspective and that of my patients is a little bit different. We're looking to ensure that if a patient goes back to the pharmacy and gets another fill of their medication that it will have the same clinical safety and effectiveness. And to be perfectly blunt, I use generic medications. I have friends who use generic medications. I have no problem with that conceptually. I want to make sure that from a patient care standpoint it will be similar. So in reality what this refers to is a patient who's on Formulation A, who goes to the pharmacy for their monthly refill, and they may either get Formulation A again, or they might get Formulation B. And the hope, the assumption in bioequivalence testing, is that one would have the confidence that Formulation A and B will be identical and work the same way.

Now, we've heard a lot of discussion about TSH as a clinical endpoint. I'm actually not going to focus on that for most of this discussion. I think it's been well demonstrated it is an important pharmacodynamic parameter, but the pharmacokinetics of bioequivalence testing are also an area that needs considerable improvement. So what we deal with levothyroxine is that of an endogenous hormone. One of the factors that hasn't been addressed today is the fact that thyroid hormone modulates its own absorption as well as its metabolic clearance. What that means, demonstrated decades ago, is that the absorption profile in a hypothyroid patient is quite different as compared with when they're euthyroid. So it is critical that thyroid hormone levels be normal when one is studying absorption and metabolic clearance.

We've had a lot of discussion about the approach to correction methodology. Even with the existing approach to baseline subtraction, as you'll see, has significant flaws that need to be addressed as well.

There are considerable sources of biological variance that come into the picture. First of all, as has been discussion, there is seasonal variation. In the summary that was published by

Andersen two years ago in the journal of *Thyroid*, it shows in that table that for the most part, the seasonal variation that's associated with T4 levels is greater than the seasonal variation that's been associated with TSH. What's more, in looking at that data, it's not quite clear that the seasonal variation has to do with the thyroid's contribution of thyroid hormone to begin with, but may also have to do with binding proteins and metabolic clearance issues that do play a role in bioavailability studies.

There is circadian variation as well, and it is true that it does seem to have a greater impact on TSH levels as compared with T4, but as has been published, and Dr. Ridgway showed you very nicely, the fluctuations diurnally in TSH do not exceed the normal ranges. So one would not be fooled into diagnosing a patient as hypo- or hyperthyroid simply because their TSH is measured at 4:00 p.m. rather than 8:00 a.m.

Another item that has not been discussed. There's considerable enterohepatic recirculation for levothyroxine. There's a considerable amount of T4 that's present in each human's gut at any given time, and as a result, the kinetics of thyroid hormone in circulation are extremely complex, and certainly do not follow the rules of simple linear kinetics in

measuring its absorption, particularly if you're following it not over a couple of hours, but over 48 hours.

There are technical issues that deal with the concentration of protein-bound substances, such as the posture of the patient, the phlebotomy conditions, whether they have a tourniquet on or off. All of that contribute to the biologic and analytical variation. There is the possibility of subject-by-formulation interaction. This is assumed not to be the case, but that is again just an assumption.

And finally, it's been commented that with levothyroxine, once the drug goes into solution, once it has dissolved, all issues of variance are really gone at that point. And that actually is not true. It was demonstrated about 35 years ago by Marguerite Hayes and colleagues, using radiotracer thyroxine in solution that there was considerable both inter- and intra-subject variation in the absorption of levothyroxine, ranging between 50 and 80 percent in euthyroid individuals, and up to 100 percent in hypothyroid. So the solution concept as outlined in this picture, may not be an applicable assumption for levothyroxine.

Finally, as has been stipulated, we're

dealing with a narrow therapeutic range drug, which adds yet another level of complexity. And therefore, we have different considerations, or certain possibilities that have to be considered in designing a bioequivalence trial specifically for levothyroxine. One has to do with the method of assessing bioequivalence. Do we deal with average or individual bioequivalence? And I'll discuss that soon. You need to consider the dose of thyroxine that's used in the absorption study. Are we talking about physiologic dosing, or pharmacologic dosing? Do we deal with single-dose absorption studies, or do we also consider repeated dose, or steady-state studies, and do we use normal volunteers, or do we use patients?

Now, all of these issues eventually percolate down to some very practical ones, which has to do with things like sample size, study duration, and the cost. It is clear that one can reduce the cost and the sample size by the use of a crossover design. However, the study duration might be considerably longer, particularly in an individual bioequivalence study. So first we'll talk about average BE, which is the methodology that's currently used, and what that relies upon is demonstrating mean bioavailabilities of two formulations being

sufficiently similar as we've discussed, not identical, but sufficiently similar. And the format for such a trial is typically a two-period randomized two-sequence study where a subject would either start on the test preparation and then switch to the reference, or vice versa.

One of the key assumptions is that within-subject variances are equal in these analyses. Now, that becomes a particular problem when we deal not with the presence of just simply one formulation and one generic equivalent, but in a drug like levothyroxine where there are multiple formulations available, the problem compounds. So in this analysis by Midha in 1998 showing that these sorts of bioequivalence criteria that are based upon average bioequivalence permit a large disparity amongst various formulations, particularly for those drugs that have a low within-subject variability like levothyroxine, and when the drug in question has a narrow therapeutic index. What that shows on this slide is that if you're just dealing with two drugs A and B being interchangeable, then as you decrease the variance in the drug absorption, you end up with a geometric mean ratio that is defined as staying -- as less than 1.2, and that's part of our criteria for

equivalence. But if you have three drugs where B is the initial branded preparation, and A and C are both declared equivalent, you can have a situation where A is equivalent to B, and B is equivalent to C, but the transitive property doesn't apply, and A is not equivalent to C. And in fact what you can see is you can have a total geometric mean ratio as you get down to low CVs that approaches 1.5. So clearly those would not be interchangeable with each other.

Now, another approach which is helpful in this sort of situation is that of individual bioequivalence. And this is a concept that the FDA itself introduced a number of years ago for consideration as a methodology for doing bioequivalence testing. What it involves is comparison of individual responses to two formulations within subjects. And it specifically applies to the question of switchability, whether you're talking about the creation of generic equivalence, or a new manufacturing methodology for the same brand of medication. And in the typical individual bioequivalence study, we address a lot of the issues that people have pointedly addressed earlier today. And that is it allows us to not only look at the variability between two preparations, but the

variability within one given preparation itself. So it typically would have a four-period two randomized sequence approach, patients starting on test, switching to reference, going back to test and then to reference, or vice versa. And the analysis of this sort of methodology allows us to estimate the within-subject as well as inter-subject variability, it allows us to analyze for subject by formulation interactions, and allows tests for both sequence, period, and carryover effects. In reality, this is what you'd be able to determine. If we have Formulation A and we want to know if they can be switched to B, certainly it allows as our average testing dose to compare A to B. But it compares that, the A to B switch, with what happens when the patient stays on Formulation A. And it's only when the variance of the A to B switch is equivalent to the variance of the A to A switch that you would declare the formulations to be bioequivalent. And I think that's very critical for the questions that have been provided for levothyroxine. Now, in this methodology, which is referred to as scaling to the reference drug, this now creates a different approach to the bioequivalence limits. Well it keeps to 90 percent confidence interval, which as FDA cites provides a 5

percent window of confidence for the patient, but it modifies the actual limits, or the goalposts, based upon the within-subject variance of the reference formulation itself. So if you are producing a reference formulation with wide variance, then it will permit the demonstration of bioequivalence of other products with similarly wide variance. If the reference formulation, however, has a very narrow variance, that becomes the same standard that any equivalent medication would have to meet in bioequivalence testing.

Single administration versus steady-state. With endogenous substances, we clearly have a problem where homeostatic equilibria affect the change in the level to minimize either increase or decrease. And so in the presence of an endogenous substance like thyroxine, it does minimize the variance in the measurements, and it reduces the sample size for bioequivalence testing, but it also turns out to maximize the likelihood of demonstrating interchangeability. This is an example, published by Marzo. If you looked at 100 microgram single-dose studies of levothyroxine, when the area under the curve variance, which is in an uncorrected model, is about 15 percent, then you can do your study with nine

subjects. However, with a simple baseline subtraction, which is what is currently used by FDA standards, it can create in the exact same study a variation of greater than 200 percent, and a sample size requirement of 2,100.

The advantage to steady-state as compared with single administration is it negates the issues of endogenous production. And as Marzo quotes, steady-state studies in instances where deficiency must be corrected, for example thyroid hormones in hypothyroidism can overcome the problem of baseline subtraction.

One can perhaps eliminate the issue of baseline subtraction by doing studies in athyreotic subjects. These are individuals who by definition have no endogenous hormone production. Now, if one uses such individuals, however, as I said, you can't leave them hypothyroid. You do have to treat them with thyroid hormone to mimic the bio-absorption characteristics of a euthyroid individual. But there are several choices, or ways one could approach it. One could use T3 or liothyronine as a way of treating the hypothyroidism and allowing the systemic T4 levels at baseline to be zero in such individuals. Now, theoretically the best way to do that would be a

patient in continuous IV liothyronine, but that's not terribly practical. But daily dosing of liothyronine can maintain the euthyroid state, admittedly with some variation during the course of the day.

The use of levothyroxine does provide us with a more stable baseline thyroid function, as well as a baseline T4, but then we have to account for it somehow in our analysis. Thyroid cancer patients therefore represent an excellent pool of individuals for such testing. The prevalence of thyroid cancer now over 300,000 in the United States, most of whom have low-risk papillary carcinoma where our data now show that greater degrees of suppression for that particular cohort is probably not of great value. And therefore, in patients who have no evidence of disease, maintaining them in a euthyroid state for purposes of bioequivalence testing would be quite ethical.

Now there have been four major bioequivalence studies that I'd like to briefly touch on that go through different methodologies. Dr. Ridgway discussed the Dong study earlier. They used two different doses of levothyroxine. There was actually one generic, it just happened to be marketed by two different companies. They used the repeated-

dose regimen, open label, four period, four sequence crossover. Twenty-four patients, those with chronically treated hypothyroidism, and they had normal TSH's at screening on these particular doses. The key things here, one is that mid-study there was a change in the lots of the medications because it took them so long to recruit individuals to that study. Secondly, they used TSH assays that are really several generations old. The inter-assay variance was 33 percent at the low end of the TSH measurements, which we would consider equivalent to a so-called first generation, as compared to the third or fourth generation assays currently available. They used a physiologic dose, and they had no washout between the periods. This is a snippet of some of the data that Dr. Ridgway showed you. Graphically, in terms of the TSH levels, although they came in normal, as he's shown you, 40 to 50 percent of the time at the end of each period of therapy their TSH's would be out of range. Not just a small difference of 1 or 2, but either going out of the normal range up or down.

Of interest as well in those data, just to go back, is they had these two doses, the 0.1 and the 0.15 milligram, but using their methodology there was no proportionality of the dose. And so the levels of

T4 that were achieved with a 0.15 milligram dose was only about 10 to 20 percent higher than that seen with the 0.1. And so there was very poor proportionality in that original uncorrected data. There was poor correlation between the uncorrected PK parameters, and the therapeutic effect of being either hypo- or hyperthyroid. There was in that study considerable TSH variability, and it was probably excessive, and it may have been in part due to the insensitive assay that was used, and the variations in drug lots throughout the study.

But there have been others that I think are more to the point. This is from Italy, two separate studies, one looking at 100 microgram tablets, and the other looking at 250 microgram tablets. And this was a within-formulation comparison, but of two different methods of preparation of the drug, of manufacturing procedure. So it was a repeated dose regimen, two period, two sequence crossover, 20 patients in each trial, again, all with normal TSH's at the outset of the study. Again, the sort of random sequence that I showed you earlier. Eight weeks of daily treatment, 1.7 percent documented frequency of missing pills. They used a far more sensitive TSH assay with a far lower variance

at the low end, and they used physiologic dosing for their bioavailability.

These are their data in the absence of baseline correction. A correction methodology was not used in this study. Like the Dong study, they only looked at the 24-hour AUCs, rather than the 48 that is now required. But they concluded in this study that test and reference were equivalent. And in this situation, TSH suggests that that really is the case. So they commented, "The values of TSH were in all cases within the normal range throughout the study period." So one can find stable long-term TSH's in such individuals, and therefore one would suggest that there was an excellent correlation between the PK bioequivalence and the therapeutic effect.

In another study from Brazil comparing two different preparations with 0.1 milligram tablets. Again, chronically hypothyroid patients, physiologic dosing. There the area under the curve for 24 hours fell into the 90 percent confidence interval of 86 to 93 percent, which would be considered bioequivalent. But one of the main differences in this uncorrected study is that you can see that the minimum and the maximum thyroid hormone concentrations on each product differed by about 1. And therefore, probably the AUC

is accounted for by the baseline change.

Finally, in a pooled analysis published last year of eight separate studies comparing various levothyroxine tablet dosage forms to liquid drug in Europe, individuals, healthy volunteers treated with a single-dose regimen, open label, two sequence crossover design. Again, just the standard random sequence. And looking at pharmacologic doses now instead of physiologic, they did the 48-hour AUC and max, and a variety of correction methodologies, including using the baseline T4 not as a subtraction but as a covariate in the analysis of variance, and a 6-week washout between the studies.

What you see here is that the residual standard deviation in the analysis of variance was quite low when you looked at the uncorrected area under the curve. When you used a baseline subtraction methodology, though, that increased by fourfold, as was theoretically proposed earlier. But if instead of subtraction you used the total T4 at baseline as a covariate in the analysis, you once again brought the variance far down, making it a tighter analysis.

What it turned out was a big part of that was probably seasonal variation in the T4 level itself, and it accounted for 10 to 15 percent of

variation in the AUC during the nine months of the study. And therefore, if you used that baseline, it corrected for the seasonal effect as well as other contributing factors of age and the volume of the thyroid gland that were found to be confounders.

So how to put all this together in an optimal study. I am a simple clinician, and so I'm doing my best to envision what would not only be pharmacokinetically valid, but also would contribute to confidence amongst physicians and patients. I think the first step is to use narrower goalposts with similar standards for test and reference products, and the use of an individual bioequivalence methodology would permit that. Second is to try to minimize the impact of endogenous substance. The use of athyreotic patients would be optimal. Steady-state measurements are both practical and reduce the impact of endogenous hormone. Physiologic dosing with the use of T4 as a covariate in the ANOVA would probably provide us with the best confidence in the analysis. And finally, and to underscore the earlier points, I think it would be extremely helpful to the clinicians and the patients in appreciating what these data would mean if TSH measurements were also incorporated to document pharmacodynamic equivalence in what I would hope would

be demonstrating pharmacokinetic equivalence. Thank you.

DR. LADENSON: Thank you very much, Dr. Sherman. The final speaker in this section is Dr. Robert Lionberger. Dr. Lionberger of FDA is going to discuss the FDA perspective on pharmacodynamic bioequivalence measures, methodological and regulatory consideration, and study design issues related to TSH and bioequivalence studies.

DR. LIONBERGER: Thank you very much. Today I'm going to talk about how FDA considers the use of TSH for bioequivalence. And to begin with, I want to remind you of what we talked about before as to what the role of a bioequivalence study is. Again, it's an in vivo confirmation of expected equivalent product performance, when we already know that the product has the same dose. We know that levothyroxine is a high-solubility drug, most products are rapidly dissolving, the absorption is limited by the permeability across the intestinal wall. We also know that there's a record of similarity of products to solution formulations. And again, the purpose of a bioequivalence study is to confirm the product performance. It's not for the bioequivalence study to be a replica or a replacement for a clinical study.

When we're talking about bioequivalence, usually the clinical safety and efficacy has already been established for the particular drugs. We're not trying to replicate that data.

And as you've seen before, this is not an unusual problem for FDA. We've had to make this decision for thousands of products. And the results of this experience are codified in the CFR. And you've already seen the quote from the regulations. And what I want to do in this talk is try to describe to you a little bit about the reasons behind why these things end up in this order, with particular reference to things you see looking at TSH and levothyroxine.

And so when we start to design a bioequivalence study, we have several choices to make. And so some of the choices that are relevant here that we've heard in some of the previous talks are whether or not we should use patients or healthy subjects, and whether the study should be a single-dose design or a steady state design. So if we just take these two degrees of freedom, there's two cases that we can knock out right away. Patients need to be treated, so we really can't use single-dose studies in patients. And we really don't want to expose healthy volunteers to steady-state long exposure to drugs that

they don't need. So those two options are out, and that really leaves steady-state studies in patients, or single-dose studies in healthy volunteers as the two choices.

And when we look at these two choices, we can really see sort of the heart of today's discussion. If you look at the first point, a steady-state study in patients, this seems very appealing because on the surface it really looks similar to what you do in the actual clinical use of the product. So on the superficial level it seems appropriate. And on the hand we have the single-dose study in healthy subjects, which is what FDA recommends to sponsors to demonstrate bioequivalence. And what we want to do today is sort of drill down and see why when we dig deeper the single-dose study is really the most appropriate way, in light of the purpose of the bioequivalence study, to demonstrate equivalent product performance.

And so first we'll look at the steady-state study, and just imagine what one might look like. So a patient comes in for a checkup, measure the TSH levels, there's no change in dose, you come back six weeks later, or whatever the duration of the study is you measure the TSH levels again. And then

you'd evaluate whether or not the TSH levels are the same. And you might do this either with a single measurement, or maybe you might measure the AUC of the TSH over the whole period.

And so this is sort of the outline of the design. One way to look deeper at this design and see some of its strengths or weaknesses is to imagine doing this study, but looking at what would happen if you used this study design to compare a product to itself. That's sort of a way to look at how good the test is, right? You know that the product is therapeutically equivalent, say different batches from the same manufacturer. And so you might refine our definition to say will the new TSH level be the same or different from the old level, even if the product and dose is the same.

Now I want to point out an important difference from this type of study and the usual therapeutic monitoring that goes on. When you evaluate a patient, you're usually checking to see if their TSH levels are within a normal range, which is not -- you're not looking to see if you get exactly the same numerical measurement. When we're looking to design a bioequivalence study, we're really looking to make a quantitative comparison that can allow us to

draw statistically significant conclusions about the differences. So we want a very strong level of precision or reproducibility in the measurement, not what you might look for in a clinical setting to find out is this patient's TSH level still under control. We want a quantitative answer, not a qualitative yes or no measurement. And because we want this qualitative statistically significant comparison, we're really worried about the sources of variability in this measurement. And we've heard lots and lots about these today already, but just to go through some of them that might come in: the time of day that you do the measurement, the compliance of patients with the product, whether or not over the duration of the study the disease is getting worse, if the patient undergoes a lifestyle change, if they undergo a diet change, if they start eating walnuts for breakfast, for example, if there's seasonal variation. How you store the product is also important. We've seen that one of the major issues with levothyroxine products was loss of potency, what we call stability. And so if the product -- and storage conditions can affect that. Also, along with that product quality issue is how old the batch is. We've seen that the potency within the product ranges from 100 percent if you have

a fresh batch, and it could fall as low as 90 percent at the end of its shelf life. And that shelf life would be different for each of the currently marketed products.

And so if we drill a little bit deeper into some of these sources of variation and sort of try to see a little bit how much they are. If we look at just time of day variation, we can see that, again as we pointed out, TSH levels within normal ranges, these are in healthy subjects, just looking sort of hourly measurements, you definitely see variations from a low of 2 to a high of 5 within the means of these data. And in this case you'd probably say if you just took those two data points, at least according to an 80 to 125 measurement of equivalence at different times of day, products might not be bioequivalent.

Again, if you do a steady-state study, you have to do the study over a long enough time for the product to maintain -- to reach a new steady state. And as we know, these products have the potential for being unstable. So if we look at just some representative data of how much product potency changes over time, we can see -- and compare that to, say, a study duration for a crossover study with just

two six-week periods -- you can imagine a study being of even longer duration -- that the product that you're using in the study might actually be changing in potency over the time of the study. And this issue is even more important when you go back and look at older studies in the literature, where the products that were used in those studies were pre-regulation by the FDA, and the shelf life, the stability overages of those products in those studies weren't very well characterized. And also the batch-to-batch variability between those manufacturing processes weren't as well characterized as they are today. So this is, again, just another concern of doing a longer term study on these products.

Also in the literature there are some of the other sources that have been measured. Subjects with sleep withdrawal, that can cause differences in TSH levels, and so if after the six weeks you happen to measure the subject at a particular time when they're getting less sleep, that could affect the variability. There are seasonal variations that have already been measured, again, that might depend on the age or the gender of the subjects as well.

So if we look at just one particular publication that measured just TSH levels over --

daily for a period of several days, you can see that for patients that were supposedly under control, you saw variation just from each day in the TSH levels. And this is consistent with what Dr. Ladenson described in his introductory talk, that currently 10 to 15 percent of the patients are either out of control right now, either high or low, 10 percent above, 10 percent below at present so that there is significant variation just from day to day within patients that are supposedly under control. And so if we think about what some of the implications of this level of variability is, what we draw from this conclusion is that based on the variability, using TSH would make it difficult to use as a precise measure of product differences. We're not very confident yet that if we did, say, a Synthroid versus Synthroid study using TSH as the bioequivalence measure, that the product would be bioequivalent to itself. Of course, that study hasn't been done, and the previous speaker indicated that he shared the understanding that that would be a valuable piece of information to have when designing a particular study.

Again, when we say the TSH levels aren't the appropriate measure for bioequivalence, this doesn't mean that it's not the appropriate measure for

clinical monitoring and treatment of patients. But again, the purpose of the clinical monitoring is to show that the patients are under control. The purpose of a bioequivalence test is to find an accurate measure of differences in product performance when it comes to the rate and extent of absorption of the drug. So again, we're not talking that TSH is not valuable for clinical use, but for use in a particular way of evaluating product formulation. And this is something that's sort of generally true, that clinical outcomes are not the most effective way to detect small differences in formulation performance. And in levothyroxine, where patients receive individually tailored therapy, and you try to do this type of comparison, each patient in your comparison would be receiving a different dose. So you'd be doing a whole bunch of different comparisons. It wouldn't be a set of patients with a 300 microgram tablet versus the 300 microgram tablet. You would have all different strengths, because you'd want to keep the patients at the appropriate level.

And so again, the goal that I think we all have, both FDA and speakers from the societies, is that we want patients to know that when they switch products the outcome will be the same as if they

didn't switch brands. Products should be -- that's what we mean when products are therapeutically equivalent. They're interchangeable. But that bioequivalence and TSH levels doesn't really appear to be the best way to achieve this particular goal, and this is primarily due to sort of the variations in the TSH levels. We've also seen evidence today of how sensitive TSH levels are to changes in T4 concentrations. But it seems also true that TSH levels would also be sensitive to other things. So you could get minor fluctuations in patient state, giving you big changes in TSH levels that wouldn't be helpful in detecting differences in formulation performance.

And so if we look for the best way to reach our desired goal, we can see we've looked and identified a lot of the potential sources of variability. And so just enumerating them again, there's differences in the variability that comes from the drug product itself, how it's manufactured, how stable it is, the amount of sleep patients are getting, the time of day products are measured, compliance, disease progression, food effects, what the patients are eating, all can contribute to the variability of the TSH levels that you might measure.

But if you look carefully at this, you'd see that almost all of these sources of variability, except for the drug product, are sources of variability that would be the same between a generic product and the reference product. And that's one of the reasons why FDA considers single-dose studies in healthy subjects the best way to focus on the drug product performance. In this type of test, we're able to remove from consideration a lot of these common sources of variability, and focus on comparing the two products directly to each other.

And again, we're looking for ways to determine equivalence in drug absorption. And I've just given an example of that in this particular slide here, showing -- this is in healthy subjects given a single dose. And we have data on the baseline level of T4 taken from the previous 24 hours, and also the baseline TSH level taken from the previous 24 hours. At Time Zero, you give the drug. Now, the absorption of the drug primarily takes place within approximately the first four hours after ingestion in terms of gastric emptying time, transit time through the small intestine. And what you see in this case is the T4 levels measured in the blood, starting at Time Zero, jump up immediately as the drug's being absorbed.

They provide the direct measurement of how fast and to what extent the drug product is providing the drug into the blood. Well, if you look at the TSH levels again in the single-dose healthy subject study, the TSH levels for those first five hours while the drug's being absorbed, they follow the baseline that you saw for the previous 24 hours. It's only in five to 10 hours after the drug's given, after it's been absorbed, after the T4 has been absorbed, metabolized to T3, interacted with the physiological control system that the body uses to maintain T4 levels that you start seeing differences in the TSH levels. And so here, this is an example of how measurements of plasma concentrations in T4 give a direct measurement of the rate and extent of absorption of the product, which is what we're focusing on.

And just to conclude by showing this list again. I hope that this talk has sort of given you an understanding of some of the reasons why we rank the different possible tests we could use for bioequivalence in this particular order. Again, the purpose of this is not to say that TSH isn't the appropriate clinical monitoring for treating patients. But because of the variability that we know is there, and because the goal of the bioequivalence testing is

really focused on formulation performance, that's why we would rank and recommend to sponsors that they do bioequivalence testing using the single-dose study measuring the direct absorption of levothyroxine in the plasma levels. Thank you very much.

DR. LADENSON: Thank you.

DR. ORLOFF: Thank you Dr. Lionberger. We have approximately an hour for public comment and questions, and panel discussion. I have on my list here one, two, three, four, five, six people. Dr. Wartofsky, I'm going to leave you to the end and you'll be the first speaker for the panel discussion. Let me call Lisa Fish from the Endocrine Society. Each person will get three minutes. I realize you've requested five, but please restrict your remarks to three minutes. The next speaker will be Howard Lando in the on-deck circle. Thank you.

DR. FISH: Thank you. I'm Dr. Lisa Fish. I'm the chief of Endocrinology at Park Nicollet Clinic, and I'm a clinical assistant professor at the University of Minnesota, which is where I did some work with Jack Oppenheimer on some of the thyroid dosing from the late 1980s that's been mentioned this morning. I should mention that I don't take any money from any company that makes thyroid preparations. I

also don't take money from the government except for Medicare reimbursement.

I'm here representing the Endocrine Society, which is the largest organization of endocrinologists, founded in 1916 with a membership of over 11,000 clinicians, researchers, and educators. We have major concerns about the safety of interchanging generic thyroid preparations, and I can't emphasize enough the concern is not with the use of generic preparations. I would be pleased to write a prescription for Mylan levothyroxine or for Sandoz levothyroxine. My problem is with patients being switched, and when my patients fill their 3-month prescriptions, the pills are changing shape each time they get a new prescription. So they can tell that the preparation has been switched.

As we heard this morning, because of the narrow therapeutic range they then call in sometimes with a variety of symptoms and need to have their thyroid levels re-checked. And this pretty much wipes out the goal of cost savings from using generics. I checked at drugstore.com for the cost of generic preparations, and Synthroid 0.125 is $40 for a 3-month supply, Levoxyl is $30, and the generic they had listed was $28. Therefore, per month, the cost

savings ranges from $0.66 to $4 per month for this dose and these three preparations, which means that when I do a single TSH level costing $74 extra from what I would normally have done, I have more than wiped out any cost savings from using the generic preparations, if we look at costs to the total healthcare system and not just pharmacy costs.

So in addition to providing sub-optimal patient care, we're creating a lack of trust in medication in patients that are on a medication for decades, and need to be taking it consistently. We're raising the risk in elderly of atrial fibrillation, and in very young people potentially causing loss of intellectual development. So we feel strongly that switching between generics for thyroid hormone is hazardous to patients, and does not result in any cost savings. Thank you.

DR. ORLOFF: Thank you. Dr. Lando. And Dr. Brent is on deck.

DR. LANDO: Hi. My name is Dr. Howard Lando, and I'm actually a practicing endocrinologist which is a bit unusual for this group, but most of the people actually see patients, and I give them all credit for it. I get to see the problems that occur because of the switches in levothyroxine preparations,

and let me just give you some clinical vignettes that I've seen.

Just so that you have a sense, I wrote a paper that I sent to you so that you would all have it, and I'm not going to go over it in my three minutes. What I am going to tell you, though, is that -- let me just give you some vignettes of some of the patients that I get to see.

Number one. First patient -- and I see about 25 to 30 patients a day, of which 40 percent of them are thyroid patients in my practice. And I see four to five days a week, day in and day out. So that sort of gives you an idea of the number of thyroid patients that I get to see, and the number of thyroid tests that I get to look at. The first patient I saw probably early last week was a patient who came to me from a primary care physician who was asking me what do I do with this patient because I cannot get their thyroid under control. Every time I come into my office, and he does a thyroid function test, at a 6-month interval when he sees them, the TSH is different. One time it's overactive, the next time it's underactive. And the first question I asked the patient was `What thyroid formulation are you taking? Are you taking Levothroid? Are you taking Levoxyl?

Are you taking a generic preparation?' The patient said, `Well, I'm taking whatever my pharmacist gives me.' And every time he goes in, every 30 days this patient goes in for another preparation of thyroid, he gets another different generic from his pharmacy. And every time he does that, his thyroid numbers change. And every time he has been changed, every six months when he goes into his primary care office, he's been given another prescription of thyroid hormone.

The second case I want to tell you about is a patient of mine who had thyroid cancer. Now, with thyroid cancer as you well know we need to keep TSH suppressed because otherwise we increase their risk of metastatic disease and progression of their disease. And this patient was well controlled on a brand of thyroid hormone. And I don't really care which brand, to be very honest about it. It doesn't matter to me. I use all the brands of thyroid hormone. It's just that I don't want my patient to switch from Brand A to Brand B. Because this patient was switched, his TSH went from where it was supposed to be to a level that was now measurable, and happened to come in with a recurrence of his thyroid cancer with lymph node metastasis. Now, can I say that it was because his TSH was elevated that he wouldn't have

had it otherwise? Absolutely not. But it certainly is something that we know is a co-carcinogen, and certainly something that we know can do it.

So what I'm trying to say to you is that think very carefully. Yes, it is the TSH that we need to measure in clinical practice. It is not T4. It is not what you're measuring for bioequivalence, or what you claim to be measuring for bioequivalence. And if we take your argument out to its extreme, what we are telling our primary care people is that, no, TSH is not what's important to measure. What's really important is T4, and we know that to be wrong. Thank you.

DR. ORLOFF: Gregory Brent, and Irwin Klein is next.

DR. BRENT: Thank you. I'm Greg Brent, a clinical endocrinologist. I'm also secretary of the ATA, and I have a lot of hats. Not as many as Dr. Weintraub, but I've had 20 years of NIH support to study basic research, thyroid hormone action and metabolism.

So sort of two points I wanted to make. First, there were comments -- in my position as secretary of the ATA, I'm the final arbiter as our public statements go out, and believe me, especially

when we get three societies together, 15,000 people, not everyone agrees with those statements, but we do have a process where we go through at least two committees, go through the council, and as Jeff knows, through all the councils. So they do reflect the best we can of the leadership of those organizations.

With my basic science hat on I'm going to raise some questions that hopefully can be provocative for the panel discussion, and it really gets to the single-dose methodology. And one thing that hasn't been discussed is a lot of recent progress in thyroid hormone metabolism, which I think is probably not taken into account. And that's, that in humans, the primary conversion of T4 to T3 is deiodinase 2. There actually have been four reports now of polymorphisms in deiodinase 2. And that gets into concepts of pharmacogenomics. This will be a perfect example where people could be profiled and predict their TSH/T4 interrelationship. There's been correlations in D2 gene polymorphisms with diabetes, with a whole series of thyroid hormone actions. Well it turns out that one of the very richest places in the body for deiodinase 2 is the pituitary gland. So in fact, rather than having to sequence everyone's deiodinase 2 gene, define the polymorphism to predict the response

to levothyroxine, we have the ability to measure their TSH. And furthermore, in the single-dose study, you dramatically in minute to minute alter deiodinase 2 activity in the tissues. So that's really -- the steady-state versus the single-dose, a major argument against the single-dose is how dramatically and rapidly you alter thyroid hormone metabolism, which is not taken into account.

And just a last sort of point on the dose, which I know was brought up as being somewhat arbitrary, I can show you a study where the individuals, one of whom was my mentor, took 3 milligrams of levothyroxine. So should we stop at 600 micrograms, 2 milligrams, 3 milligrams? And I think that what we've seen as pointed out, some of the deficits of the single-dose study. Thank you very much.

DR. ORLOFF: Thank you. Irwin Klein. And then Sally Schimelpfenig, do you want to speak next?

DR. KLEIN: Good afternoon. My name is Irwin Klein. I'm a professor of medicine and cell biology at NYU School of Medicine, and chief of the division of endocrinology at North Shore University Hospital.

I'd like to direct my comments as to what

is the best way to assure the stability of the treatment of our patients with hypothyroidism. My career has been directed at the study of the thyroid hormone effects on the heart. About three years ago I had the privilege to edit this issue of the journal of *Thyroid*, directed solely to the cardiac effects of thyroid hormone.

We know that the heart is one of the most sensitive organs in response to thyroid hormone action. In my annual care of thousands of patients with thyroid disease, our standard of care evaluation is to study blood pressure, pulse, the overall clinical assessment of patients, and to confirm that assessment with measurements of TSH done on a single annual basis. That constitutes the standard of care. We've heard, however, that it's possible for the dose of T4 to be changed as much as 12 to 12.5 percent as the result of the switch to a generic preparation, either on an authorized or unauthorized basis. I can tell you from my research work, and my review of the literature, that that can produce sub-clinical hyperthyroidism in a significant number of patients. And what do we mean by that? That's a fallen TSH with the normal measure of total T4, free T4, and total T3. So in fact, we cannot diagnose sub-clinical

hyperthyroidism purely based upon a T4 measure. And in fact, the heart does not respond to T4. T4 does not act directly on the heart. So in the face of no change in serum T4, with a fall in serum TSH, we know that a significant percentage of those patients are at risk for atrial fibrillation.

Atrial fibrillation develops as an acute event. There is no time limit placed upon the period of time when that may occur. It can occur after days, weeks, months, or years. Perhaps no better example of that is the fact that our 41st President presented with the first manifestation of his hyperthyroidism as a result of atrial fibrillation.

So what then are we to conclude from these observations? The current guidelines for bioequivalence do not evaluate the therapeutic equivalence of thyroid hormone at the level of the heart. To assure both efficacy and safety for our patients, TSH measurements must be part of our evaluation, because otherwise it will be very hard to justify to our patients, especially that growing population of older patients who present to us for the first time in atrial fibrillation as a result of the change in their medication.

DR. ORLOFF: Thank you very much. If

there are no other speakers from the audience, Dr. Wartofksy, do you have a comment or a question for the panel? You can stay at your seat if you'd like. It's up to you.

DR. WARTOFSKY: I wanted to respond, Dr. Orloff, to a couple of comments made by other speakers, if I might. One, I'd like to agree with Dr. Lando in terms of prescription of products. The point is it doesn't matter whether it's branded or generic as long as it's consistent. And the problem I get into that I'm going to allude in my talk with switching is when patients are switched not simply from brand to brand, or brand to generic, but from generic to generic. Because the generics are different. So that once that switch is made to generic, we as clinicians lose all knowledge and control of what our patients are on.

In regard to Dr. Weintraub's comments about why T4 might be better than TSH, Dr. Ridgway outlined that. But all of the problems that Dr. Weintraub alluded to of TSH do not apply to when we're testing for bioequivalence. We're testing under the guidelines of the FDA, of normal volunteers, euthyroid, et cetera, and not the euthyroid sick when T4 is also abnormal, or other problems, when TSH is

altered T4 is also altered. His issue about sub-clinical disease taking years to develop, Dr. Ridgway addressed, but also when we're talking about children, infants who are either under or over dosed, we can't wait years for effects. When we're talking about the elderly who are vulnerable to atrial fibrillation, we're not talking about years for that problem to arise, or the pregnant woman who can have abnormalities in the fetal brain development within weeks and months, not years, for problems to develop.

In regard to Dr. Lurie's comments, Public Citizen, very admirable, very passionate, but I'm afraid often wrong in some distorted comments. Although the three societies did fund the consensus panel that was published in JAMA, the three societies did not agree with the conclusions of the consensus panel, and that has been published, which he failed to cite, in all three major journals of the three societies. But the societies did not suppress the opinions of the consensus panel. So while admirable and well-meaning, physicians and Public Citizen who have little or no endocrine training are coming against the thousands of endocrinologists in the professional organizations who feel otherwise. And Public Citizen, I'm afraid, is the one that is stuck

on Groundhog Day.

DR. ORLOFF: Thank you. Maybe I could just make a point of clarification based upon the definitions that are being bandied about today, and then ask a question which I hope will stimulate some discussion.

In my career, not as long as many of the people seated on this panel, but as long as I've been an endocrinologist and a physician, up until 1997 there were no generic levothyroxine products. We need to be clear that although the nomenclature in the endocrine and thyroid field was brand name versus generic, and although the rule of thumb was that brand name was good and generic was inferior, brand name was a known entity, generic was an unknown entity, you must understand, everyone in this room, that it is only subsequent to the approval of the first new drug application for a levothyroxine sodium product in 2001 that we could possibly have generics. And as you've heard, and as we'll discuss further, the generic products that we have on the market today are -- they're not generic because they say "levothyroxine" on them. They are generic because they are deemed therapeutically equivalent to a reference product. And let me just say one more time, I know it's been

said many times, but that determination of therapeutic equivalence begins with the determination that they are all of equal potency. And the second part of that determination is that they are all readily dissolvable and indeed, they all dissolve, in vitro at least, to 100 percent, and are presumed to do so in vivo. And then, as follow-up confirmation, in order to be sure that we haven't missed anything, say for example that there's something weird, a weird excipient that got in there by mistake, or that we didn't previously understand might interact with the absorption of levothyroxine, they are tested in a bioequivalence study. And that bioequivalence study is simply a measure of the degree to which the content levothyroxine of the product is available for absorption through the intestinal wall. Period. The degree to which it is available for absorption.

So differences observed in bioequivalence studies can be true differences, they can be related to true differences in the availability of the levothyroxine in the product, they can be related to differences in the potency of the two products being tested, because although we use a quantitative analysis, or the companies use a quantitative analysis, i.e., HPLC, to determine the potency of the

products that they're going to use in the bioequivalence study, it turns out because the test, or the generic company has to go buy it off the shelf that many times they cannot get a product that has precisely equal levothyroxine content as their product. So there's always a difference at baseline. There is also the potential for decay in potency over the 35 days. And then the final thing that can contribute to an observed difference in a bioequivalence study, or confirmatory demonstration, is intra-subject and inter-subject variability in absorption.

And I should add one more thing, which is that these studies are not powered as hypothesis tests. They are of fixed, to some extent arbitrary sizes. You heard one generic sponsor, I believe it was Mylan, make note of the fact that they generally use larger numbers of patients in their bioequivalence study. The reason there is a purely statistical one. It narrows the confidence around the mean observed difference.

Anyway, let me follow that, and if I might ask a question for discussion. I think we would all agree that the ideal levothyroxine sodium product is one that is quantitative in its potency, that is

stable, optimally stable, over its shelf life. Ideally we would like it to retain 100 percent of its drug content, active drug content, from release and shipment from the factory to the last pill the patient takes at the last day of its shelf life. So we would like it to be optimally stable.

And then finally, we would like all of that levothyroxine that's in the pill to be bioavailable. That is to say we don't want a pill that doesn't dissolve completely. We don't want a pill that turns into a slurry as opposed to a solution in your stomach. We want every molecule of levothyroxine to be freely in solution, in the gastric and intestinal aqueous contents. That is the ideal formulation. Parenthetically, we believe that all of these products adhere to essentially -- to acceptable standards in that regard, although there will be discussion, as I think you already realize, that there are differences in the rate at which different levothyroxine products lose their active drug content.

But I guess what I want to know is there has been a focus all day today on the observed difference between the Abbott product in the bioequivalence studies, in terms of its bioavailability, and some of the products to which

it's been compared, which if anything would suggest that the levothyroxine content of the Abbott formulation is not fully bioavailable. And I'm curious whether anyone on the panel would like to address what might be going on there, or whether anybody from Abbott would like to address what's going on there. Because, as I said, the most -- the best product we could imagine is one that has fully bioavailable levothyroxine content. If anything, that product, based upon the societies' reads of the data, does not have fully available drug content. Are the differences we're seeing there related to intra- and inter-subject variability? Are they related to differences in potency at baseline? Are they related to differential loss of potency over the 35 days between Period 1 and Period 2? Question for discussion.

DR. RIDGWAY: Well, I didn't mention the Abbott product, and I wasn't talking about Synthroid. I was talking about the switching between one drug and another. And you just asked a series of questions about what could account for the variability. And so I would like to ask the FDA exactly --

DR. ORLOFF: No fair asking a question after a question.

DR. RIDGWAY: -- exactly why the FDA won't do the study to find out about that variability, and then to incorporate it into the model, what the results are. What is the fear of doing that? And this idea that there's too much variability in TSH is just not correct. And we ought to test that. Why are we afraid of getting the data? FDA wants to find this business about dissolution, and about performance, and about bioavailability, but if they want to do that, and then they want to recommend that you can switch those two, you ought to do the study on the patients.

DR. ORLOFF: Well, let me -- honestly, I would like to hear some discussion of what is the basis for the difference in bioavailability. But we can address the question of who is going to do a study to affirm FDA's methods or not. I don't think FDA is going to do it. But I guess what we need to understand around the table here is if you put the same amount of levothyroxine into one pill as another pill, and let's take it on faith that an HPLC is a highly precise assay. So the potency assays for these products are to be relied upon. If you put the same amount of active ingredient into two different pills by the same manufacturer or by different manufacturers, what can account for the differences in

the amount that gets absorbed out of that pill?

DR. HENNESSEY: I'd just make a comment that obviously with a 5 percent molar ratio that's required for the bioequivalence studies, that it's supposed to be measuring apples to apples, and comparing apples to apples, at least with pharmaceutical equivalence. So in my mind the only difference can be in the constitution of the excipients, and how the dissolution occurs amongst the pills. And there may be differences in bioavailability, but that's really what it is, differences in bioavailability. And we aren't talking about a pill that might have a different bioavailability not being able to deliver a specific amount of thyroid hormone on a consistent basis. We're simply talking about differences between preparations that then if substituted might lead to a change in the overall thyroid function assessment.

DR. ORLOFF: And what makes you think that then when we actually have an observation in a bioequivalence study, a confirmatory study after quantitative assay of drug content and dissolution between, for example, Unithroid and Synthroid, also on Dr. Davit's slide, where the ratio of the AUCs 0 to 48 is something like 1.03, do you think that those two

are also not therapeutically equivalent? What's your concern there?

DR. HENNESSEY: All I can say is that the two observations that I saw were 12.5 percent difference and 9 percent difference in the AB2 rated products, and potentially the third pairing could be. But a clinician, of course, is going to be measuring a TSH in a patient, and that could turn out to show something different.

DR. WARTOFSKY: Dr. Orloff, I think what our three societies are after is for the FDA to tighten the goalposts, to have more stringent criteria. And if Abbott's product is not meeting 100 percent content, then it's declared bio-inequivalent. If you tighten the goalposts and have more rigid standards that everyone has to meet, we'll be happy. That's for all the brands, whether we call them generics or brands, that's for everyone.

DR. LADENSON: Dr. Conner, did you have a comment? I missed you reaching for the microphone there.

DR. CONNER: No, I've gone on to another topic.

DR. LADENSON: All right.

DR. KLEIN: Coming back to your question

directly, because I think it is an important observation. Three, perhaps four of the agency spokespeople have referred to the fact that levothyroxine sodium is freely soluble. Two questions. What's the basis for that conclusion, and in fact, what is the solubility of levothyroxine sodium? Because in fact, if we're dealing with solubility issues, and it's not freely soluble, many of the assumptions in your bioavailability studies are not correct.

DR. ORLOFF: Dr. Malinowski.

DR. MALINOWSKI: I think I can answer that. And it's something that hasn't come up yet today, and there is something called a Biopharmaceutics Classification System, which has been developed by FDA, and has been implemented for classifying drugs as highly soluble, or low solubility, highly permeable, and low permeability. And that's been implemented to the extent for highly soluble, highly permeable drugs. Bioequivalence studies are not needed because there are thought to be no concerns about bioavailability.

So getting specific to your question, our laboratory has tested the solubility of levothyroxine --

DR. ORLOFF: Please speak into your microphone. Put it closer to you.

DR. MALINOWSKI: Our laboratory has tested levothyroxine specifically to your question, and has determined that it is high-solubility, and that it would take only five milliliters to dissolve the dose, the highest 300 microgram dose of levothyroxine. All I'm reporting is what our laboratory has done, and that is real data that can be relied on.

DR. LADENSON: Yes, sir, would you come to the microphone, please?

DR. JERUSSI: My name is Bob Jerussi. I can speak loud enough. Levothyroxine sodium is very soluble, when it hits the stomach, it no longer has the sodium salt. It's levothyroxine. What is the solubility of levothyroxine?

DR. LADENSON: Dr. Malinowski?

DR. MALINOWSKI: The data I referred to, done by our laboratory, and for the Biopharmaceutics Classification System, has to be conducted over a range of physiologic pH's. So that was accounted for.

DR. LADENSON: Yes, Dr. Landschulz.

DR. LANDSCHULZ: I'm Bill Landschulz. I'm from Abbott. There seems to be some controversy still here about solubility, etcetera, about levothyroxine

products, but what I'd like to say is that we clearly -- Abbott product clearly meets all specifications, quality specifications that have been instituted by the NDAs. We applaud that. And to amplify Dr. Wartofksy's comments is that I think that if there is an issue, that we should be looking at the 80 to 125 boundaries, and getting a better understanding of why we believe that that is acceptable for this narrow therapeutic index product would be I think very useful.

DR. LADENSON: I'd like to comment if I could, Dr. Orloff, and it really follows up on that precise point. What bioequivalence testing is all about is the issue of rate and extent of absorption. And although these compounds differ from one another, that's precisely the reason that that is part of the FDA's criteria for equivalence of these drugs. And what the clinician has to cope with, as you've heard again and again from clinicians, is the fact that the patient is on one approved drug and switched to another, where the FDA's own current bioequivalence standards show a difference that FDA itself has recognized are outside of the boundaries of acceptable changes in dose. And changes in dose that have potential clinical consequences.

So I think we've got to see this promise that compounds that differ by 9 percent or more not being approved. We've got to see that promise honored. And that's what our societies are concerned about, and it is bioequivalence testing that is telling us that that promise has not been fully fulfilled.

DR. ORLOFF: Let me just respond to that to clarify. There is nobody who's worked on this at FDA who is not absolutely certain that precision in the dosing of levothyroxine is very important to appropriate management of patients requiring levothyroxine therapy for its various indications. Precision in dosing. Precision in dosing is not -- precision in dosing starts with the potency of the tablet, the amount of drug in the tablet, and then it goes to certain qualities of the tablet that have been discussed, that are assessed in an ongoing fashion during continued manufacture of the tablet; that is to say, dissolution profiling. And it is confirmed by the bioequivalence tests.

But I think there is a confusion here. The societies have taken a mean -- any of the mean differences that are observed in these confirmatory in vivo tests. These are tests of the product in an

imperfect animal. It's not being given intravenously. It's not being given intramuscularly. It's being given orally. These are used as confirmatory tests for our assurance that there isn't something crazy going on that we were not otherwise suspecting. But the societies have looked at these observed differences in the means, or indeed at the outer limits of the confidence intervals as representing a possible difference in the quantitative, essentially, delivery of drug.

What we have talked about in the past with regard to precision in dosing, and the necessity to adhere to less than 9 percent differences relates to product potency. We do not believe that the bioequivalence test is a quantitative measure of product potency. On that we don't -- in a sense, we don't disagree with you, but you believe that the only way to know if two products are the same is to study them out for six weeks in a crossover design to look at TSH maintenance in an athyreotic patient. We would say, and we've said it many times, that our scientific principles, and our drug manufacturing principles, and biopharmaceutic principles tell us a priori that these drugs are essentially all the same, even before the bioequivalence test. But we do require a

bioequivalence test as a formal demonstration in order for a regulatory declaration of therapeutic equivalence.

DR. WARTOFSKY: Could I comment, Dr. Orloff? I think, and correct me if I'm wrong, that one of the major goals of the FDA is to ensure safety and efficacy of pharmaceutical products. And that first step you allude to of precision in dosing doesn't do it. What we're telling you is it doesn't do it. It assesses bioequivalence, and you say the precision in dosing is confirmed by the bioequivalence testing. But it's not confirmed clinically. We're telling you that we're not seeing that confirmation in our patients. Therefore, something has to change in that bioequivalence testing to be true bioequivalence testing.

DR. ORLOFF: Well, I guess I think what's going to come out of today's conversation is that a confirmatory or refutatory study, and I believe it would be on the part of the societies, because I don't think it's going to come from industry, such a study to TSH endpoint is going to be required to resolve this in your minds. In our minds, we believe that our standards are scientifically based and reliable.

DR. LADENSON: You know, as we were just

talking, you were talking about in vivo experiments in imperfect subjects, that's what a clinician does all day is deal with, you know, the reality of where the rubber meets the road. When a patient swallows a pill, and what the clinical and biochemical outcome is. And that's why I think we're very concerned, based upon the bioequivalence standard that those in vivo experiments in imperfect models, the average Joe taking thyroxine is telling us that using properly statistically determined experiments, that we're seeing differences of as much as 22 percent. And I think, you know, this could boil down to something as simple on the bioequivalence side as just a willingness to look at this again and narrow the goalposts, and knock that kind of difference out of the clinician and the patient's life.

DR. ORLOFF: Well, the goalposts could be narrowed simply by increasing the size of the studies. Remember, the goalposts are -- virtually all of the tests for both bioequivalence between products and dose proportionality within products, which is another critical aspect of the utility of individual levothyroxine products that you know and I know when I treat a patient, or when I up-titrate a patient from 100 to 112 micrograms, that there is an additional 12

percent, not 12 micrograms, there's an additional 12 percent of available -- of bioavailable levothyroxine sodium in that pill. The studies that we've done to establish dose proportionality and bioequivalence between products all fall -- the 90 percent confidence intervals all fall well within our goalposts, as you suggest. But narrowing the goalposts, or narrowing our confidence is really a matter of doing larger studies. That's not necessarily going to change the variation you're going to see around unity in the observed means from one study to the next.

And I just want to say, Dr. Wartofksy and Dr. Ladenson, please, no one in this room, nor should the societies believe that we have anything but the best interests of patients in mind. I too treat patients with thyroid disease. I have their best interests in mind. We do not have clinical trial data, or even particularly good observational evidence, to the extent that it would be reliable at all, that there are any problems out there. We have anecdotes that give you concern, but your concern is based upon an a priori failure to accept the standard because, we believe, of a misunderstanding of actually the interpretation of that bioequivalence exercise.

DR. LADENSON: It looked like Dr.

Malinowski, and Dr. Sherman, and Dr. Garber. Could we, Dr. Malinowski?

DR. MALINOWSKI: Can I ask a question?

DR. LADENSON: Sure.

DR. MALINOWSKI: I'm trying to understand better your discomfort with what we've done, and I'd like to have you comment on something, and it may not be a yes/no, black and white answer and so forth, but I'd like to hear from you. If instead of tablets that are marketed, levothyroxine was marketed as a solution, as an oral solution, how would that -- would that give you more comfort, or would you still see issues? Could someone comment on that?

DR. WARTOFSKY: I think if the -- and the solution was being marketed by a number of different companies. If the solutions were the same, the same solvent, the same everything, and there were both your bioequivalence testing and our clinical data that would confirm that they were the same, that we didn't see the major changes we're seeing now when preparations are switched, liquid would be fine. Certainly.

DR. MALINOWSKI: Well, thanks for that comment because that does help me understand that particularly your issue is with what we consider small

differences among the various tablets that are marketed.

DR. WARTOFSKY: Differences perhaps of excipients, whatever, the compacting, whatever the differences are that translate into our seeing different -- clinical differences. We seem to be talking about two different things. The FDA is talking about their precision dosing, the bioequivalence testing, and what we're saying is that does not translate on the clinical side to true therapeutic equivalence. And the issues you raise about all of the other variabilities in your talk, all true. But you heard this morning several speakers say when you add one more variable, you're just compounding the variables. So that is really not an argument that holds a lot of water. Yes, there are variations, and as you said, they apply both to branded and generic, and those are washes. But when we're getting differences in the products because the testing is not sufficiently rigorous, that's where we as clinicians have problems.

DR. LADENSON: Dr. Sherman?

DR. SHERMAN: I have two questions, perhaps for clarification. And the first is it's my understanding that the requirements in the dose

proportionality studies did not involve corrected thyroxine concentrations. Is that still the case? And therefore, are the dose proportionality studies that have been used for all of the approved products actually represent the previously flawed approach, or the at least adopted baseline subtraction? And then I'll have a second question.

DR. MALINOWSKI: The only dosage form proportionality, I call it dosage strength, in the equivalence study were in the NDAs. So in the ANDAs, all the other strengths are waived. Correct? So then focusing on your question, those studies are in the NDAs, and what I presented, as was submitted by each of the NDAs, which is uncorrected data.

DR. SHERMAN: So the proposition that the dose proportionality studies of the products themselves demonstrate their appropriate potency is based on the older methodology?

DR. MALINOWSKI: We answered that question in one of the previous go-arounds on this, that one of our reviewers re-did some of the data that was submitted in the NDAs, made corrections, and it didn't make any difference. The point I was making this morning in both of those studies, if you look you can see, it starts at a value like 7, and that's baseline.

And there is a rapid increase for solution, there is a rapid increase for the tablet. So those studies were not strictly bioequivalence studies, but I think they were the initial basis for us getting a lot of confidence that you can get tablets that have very good absorption.

DR. SHERMAN: And then the second question. When one of my family members who has hypertension goes and gets a refill on their antihypertensive, and they receive a generic product, there's no instruction in the product insert material that says you better go back to your doctor's office and get your blood pressure checked because you're on a different formulation. If FDA is confident in the true nature of equivalence amongst thyroxine preparations, then why is it in the product inserts that it says if there is a change in formulation the patient should have a TSH level checked, I think six to eight weeks later? It would appear to be inconsistent.

DR. ORLOFF: Well, that is inconsistent, and that's I assume because we have not amended those labels. But you're absolutely right. There is no -- we do not believe there's a basis to re-check and re-titrate when switching to a therapeutically equivalent

product.

DR. LADENSON: Dr. Landschulz? Oh, Dr. Garber.

DR. GARBER: I'm not sure who to point fingers at because we know the FDA is at least responsible for the safety of our citizens, and at least from a medical point of view. But you basically -- and putting aside what I think is, you know, we could argue all day long about whether 12 percent difference should be the difference or not -- but by your own admission you haven't taken every product, that is every brand product, and every generic product, and made any claim that they're all equivalent across the board. Correct? So what you've done is set up a system that's so complex that the typical pharmacist, unless he has a special interest in this, who's willing to go to a grid and know what's substitutable, couldn't even make the right -- would flunk any kind of quiz on the spot about what's a fair switch.

So it's one thing to have a concept that you have some equivalence, and a generic might be equivalent to a brand product, but when you have a surfeit of options out there, in a sense you're endangering the public by making them vulnerable to

what will never be a totally effective education program for pharmacists, won't be a comprehensive patient education program for patients, and physicians as well.

So unless you told somebody like me that you've narrowed the window, and tested everyone across the board so we knew -- we know that A is equivalent to B, B's equivalent to C, but A isn't C, what happens when you get to F, G, H, I, J, and K? So I think as much as there may be some rigor in how you've established the early phases of the comparison, it's not being done across the board and it really sets us up for everything I think we ultimately, even though it doesn't sound like we agree about too much, at the end of the day we'd probably agree is a difference.

DR. ORLOFF: Well, we understand your point. It's worth, I think, clarifying for your sake, not that it necessarily helps your perception of the situation, or in fact the reality of the situation, but we can't mandate that different drug companies conduct studies against other products in order to establish therapeutic equivalence. Indeed, as you can imagine, for certain competitors in the marketplace there is in fact a disincentive to conduct such studies. So it's the job of the little guys to define

themselves as therapeutically equivalent to the big guys, but as you suggest, the matrix gets pretty complicated.

DR. GARBER: So, could I just briefly respond to that? You would think as a taxpaying citizen that I would like to think that the FDA was not only empowered, that it would think of that and protect me by coming up with a mechanism to assure that happened. Otherwise, you basically are setting up a system, just like if we set up a therapeutic plan for any patient we took care of that was unworkable and unexecutable, we're kidding ourselves. So perhaps we can work on that together. Thanks.

DR. LADENSON: Yes.

DR. LURIE: I guess I just, responding to the last point, as I raised in my comments, there is indeed this grid, and it has many, many holes in it, and I've suggested that, you know, responsible pharmaceutical companies might be interested in filling in the grid for us. But if not, the government has a role I think in trying to fill in the grid so things get simpler. But regardless of that, the FDA is being very clear that the only issues of substitutability are between those particular pairs that have been compared. So the issue of narrowing

the goalposts is a completely separate matter from the matter of the grid. And the grid, as it currently stands, is really a matter of communication with pharmacists, and I think is an area in which the FDA could be doing more.

I will point out, though, that when it comes to the matter of filling in the grid, yes, it's absolutely right that the logical way to do it would be to have the reference-listed drug be one of the better selling drugs if the object from a public health point of view would be to take people off those more expensive but we hope bioequivalent formulations onto less expensive but equally active ones. But in fact what happened is that Abbott made an effort to have itself de-listed as a reference-listed drug so that it would be difficult for any of the small guys to be declared bioequivalent to them. So in that we see the true motivation.

DR. LADENSON: If there are no more comments at this time I think we'll move ahead with the next presentation by Dr. Wartofsky. And Dr. Wartofsky, who is professor of medicine at the Uniformed Services University of the Health Sciences is going to speak on society concerns regarding current U.S. prescribing and dispensing practices.

DR. WARTOFSKY: I feel like I've already spoken.

DR. LADENSON: And president-elect of the Endocrine Society as well.

DR. WARTOFSKY: I don't have to belabor the definition of narrow therapeutic range or index drugs. That's been commented on several times, and would point out at the bottom of the slide the similarities to warfarin, or Coumaden, Digitalis, and phenytoin or Dilantin, how important it is to carefully control the therapeutic range of these drugs, which we do by measuring their levels. My topic is switching of thyroxine products. And to give you a little background, the switching is dependent on where you live. Often we ask physician prescribers are not informed of a switch when it occurs unless that's mandated by regulations in the state, and often not even then. We find that pharmacies are not honoring the brand or product that we write for, even when writing "brand necessary" or other admonitions to do so. Rather, products are commonly switched, and they're switched often at the time of being refilled. This can cause many telephone calls between pharmacists and prescribers, and faxes, and creates a lot of paperwork and business at both ends.

Some of the issues are that branded is frequently switched for generic. I believe personally that pharmacies have a profit motive in doing so. The switch becomes confusing to patients. Approximately 18 to 20 percent of patients get confused, stop their medication for some time, until they can contact their physician and clarify the issue. When polled, patients often do not know what product they are taking.

In terms of state regulations, most of the states are what we call Orange Book states, where the pharmacist is permitted to switch, to interchange products that are declared therapeutically equivalent by the Orange Book. Then there are individual determination states that work under a slightly different system, and Virginia is our local state that has a positive formulary, and only products on the formulary may be substituted. And finally, there are so-called professional judgment states where the pharmacist can use his or her professional judgment to make a switch. That's shown here with the Orange Book states in pink, you can see, covering most of the country, including Maryland, and D.C., and Virginia there being a formulary state.

What is the impact on physicians? This

causes our patients to come back again for reevaluation, for TSH testing. We need to justify the payment for that TSH testing. The patients, whether the symptoms are due to the switch or not due to the switch, the occasion of the switch is the stimulus for them to complain about symptoms which then require investigation and evaluation. And again, more telephone calls, more faxes. The impact on the patients themselves, they don't feel well whether due to the switch or not. The inconvenience of making these additional visits, the cost when not fully reimbursed, when they have co-pays for the extra TSH testing, as well as the risk for adverse effects of either too much or too little levothyroxine.

So the question is how can we as clinicians control our patients' TSH levels, maintain them where we want them, either in the therapeutic range, in the euthyroid range, or for cancer patients in the suppressed range. How do we keep them where we want them when the pharmacist keeps switching so-called equivalent thyroxine preparations? The FDA guidance in 2000 stated that substitution could lead to sub-optimal responses, and even hypothyroidism, or hyperthyroidism with its toxic manifestations, and there was a risk in patients with underlying heart

disease that a small increase in dose could be hazardous.

Indeed, when preparations are switched, there are three questions we could ask. Will we get reimbursement for the repeat TSH testing? What is the impact on the test? What will that lead to? And how often, actually, is re-testing done in the physician community? And re-titration of the thyroxine dose as a consequence of the re-testing. In the Federal Register, in regard to Medicare reimbursement, it stated that it would be covered or reimbursed up to twice a year in stable patients, but it could be reasonable in other occasions where it could be clinically justified.

In a Pharmetrics study looking at approximately 36,000 patients who were on stable thyroxine dosage and given new thyroxine prescriptions, 70 percent of them were not re-tested within 90 days as recommended by the practice guidelines of the American Thyroid Association, and the American Association of Clinical Endocrinology, even though Dr. Orloff surprised me a few moments ago by stating that he thought this could be taken off the label, that re-testing was not necessary. In 30 percent, re-testing was done before and at three

months after, and what did they find? They found that prior to the switch in preparation the TSH was 2.39. After, it went up approximately 1 milliunit per liter. In fact, almost half of the patients had a change of greater than 1 milliunit per liter, 25 percent greater than 2 milliunits per liter, for a mean increase of about 1. Indeed, as Dr. Ridgway showed you, the Andersen study, where the variation in individuals was followed over a year, this change is greater than the variation in normal euthyroid individuals. And in fact, the National Academy of Clinical Biochemistry has published that a change of greater than 0.75 milliunits per liter is a clinically significant change. These are all changes occurring after switching.

Stelfox looked at a similar issue at the Peter Bent Brigham Hospital, 400 outpatients on thyroxine, looking at whether they received the recommended monitoring. A little more than half were counseled in terms of recommended follow-up and TSH testing after a change, and there were adverse drug events reported more commonly in those patients who were not monitored, who did not get a TSH re-measured. And there were adverse events on both ends, both the hyper end, atrial fibrillation, tachycardia, other

symptoms, as well as the hypo end.

So what is the cost of switching? There's the cost of the drugs, the impact of the loss of the euthyroid state, increased costs for TSH, for more visits to the physician, for the evaluation and assessment of symptoms that may or may not be thyroid-related, the impact on job productivity, loss of work, quality of life, and other costs. You've seen this slide before of the multiple dosage strengths of the levothyroxine preparations, and the fact that the Blakesley study demonstrating the inability to distinguish a 12.5 percent dose difference. And our belief that these small differences have a significant impact on patient safety and the efficacy of therapy.

So what are the consequences of switching, of interchange and substitution? Dr. Ladenson showed this slide, a similar slide of the vulnerable populations, the populations of patients that we worry most about. The older patients at risk of heart disease and osteoporosis, the pregnant patients, and our thyroid cancer patients that have to be very carefully controlled in regard to their desired TSH range. And perhaps even more importantly, children, particularly children in the growth ranges of their early years.

What are the adverse consequences of a potential switch and a change in potency in these populations? These are data from the National Cooperative Thyroid Cancer group of over 1,500 patients showing the difference in survival, in death rates, when the TSH was well controlled, low/normal to normal to elevated, poorly controlled. Highly statistically significant differences on mortality, on death rates, related to how well the TSH is controlled.

That's thyroid cancer. What about miscarriage, fetal demise? This is data from Allan, the State of Maine screening study looking at the fetal wastage rate, whether the TSH was above 6 or less than 6. And I believe a normal range for TSH is somewhere up to about 2.5 or perhaps 3. And here the cutoff was a very generous 6. And you can see a fourfold greater risk of fetal death with a higher TSH. We know that there is an increased demand for thyroid hormone in pregnancy, on average, approximately 50 micrograms per day, and yet many of our pregnant patients are not tested, are not measured, dosages are not adjusted, and when we're dealing with switches that can include 12 to 20, or 25 percent differences, that can lead to increases in TSH

like this, and sub-clinical hypothyroidism, miscarriage, fetal death.

In addition to fetal death, the issue of fetal brain development that was alluded to briefly earlier this morning. The study of Haddow in the New England Journal where the offspring of women with sub-clinical hypothyroidism were evaluated between ages 7 and 9 with IQ testing, and the frequency of IQs less than 85, 20 percent compared to 5 percent in the controls. Fourfold increase with failure to treat sub-clinical hypothyroidism in the mothers.

Recently, and this next couple of slides are not in your handout. This is fresh data of this week. The ATA and AACE sent out a quick snap poll questionnaire to its members this week with a couple of questions. Pharmacists substitute my prescriptions for a specific brand of LT4, even when instructed to dispense as written. How often does this happen? The second question, when you have patients under consistent good control on a specific brand, and then they present with symptoms of either too much thyroid hormone or too little thyroid hormone, how often do you find the explanation being a switch? And here are the responses to the first question. Pharmacists switch my prescription where I state a specific

product rarely, 30 percent of the time, often 62 percent of the time, and two-thirds of those "often" on a daily or weekly basis, clinicians writing prescriptions. The second question, patients under consistent control, and then you find that they've gone out of control. How often do you find that they were switched to a different brand or a generic? Twenty-five percent rarely. This is about one thousand respondents. Seventy-three percent quite often, and again over half of those on a daily or weekly basis. This is happening to us every day. I see patients. I get these calls every day, from patients, from pharmacists.

We asked two more questions. Do you support more stringent bioequivalence standards for levothyroxine product? Do you want the so-called goalposts narrowed? Ninety-five percent yes, 1,013 respondents. The last question, do you support stronger policies that would limit a pharmacist's ability to override physician orders for a specific product? Again, 96 percent yes.

So, what I conclude. We've heard thyroxine is the synthetic version of an endogenous hormone, and it has a narrow therapeutic index, like Coumaden, or warfarin, like Digoxin, like Dilantin.

Physicians carefully titrate thyroxine products, measuring TSH as their guide for the therapeutic equivalence of those products. Very small differences in dose or in product content result in significant changes in TSH. And because of the risks that are associated with these changes, with minor degrees of over-treatment or under-treatment, we are concerned that we are putting our patients at risk. Switching after a patient is stabilized causes us to lose our control of the desired patient's level of thyroid function. We see little evidence, despite the FDA's position on product dosing, bioequivalence testing. We see little evidence of true therapeutic equivalence of levothyroxine products. Switching increases the chance of adverse outcomes. I cite the Stelfox data. It increases physician and pharmacist workload without economic benefit. In fact, the increased cost mentioned by Dr. Fisher earlier on TSH testing. We note that the large pharmacy chains encourage or even mandate switching for a profit motive, and I would repeat what I said from the panel desk, that one generic levothyroxine does not equal another, and therein lies one of our major problems when our patients get that first generic du jour from the pharmacist. The next 30 days or 90 days, it will be a

different one, and the likelihood of re-testing and re-titration at that time is much less.

So finally, we need better methods to determine equivalence of narrow therapeutic index drugs like thyroxine to minimize the impact of switching. I don't believe that current FDA recommendations for bioequivalence are sufficiently sensitive to detect the small differences in products that are clinically important to us. The impact of switching is not being routinely detected by monitoring. Again, the Stelfox data, as well as our own empiric experience. Small differences are indeed important. They have significant clinical impact on safety, and patient wellbeing, and risk of progression of disease.

As I think almost every physician who got up and spoke here today expressed a sense of frustration at the current situation as being unnecessarily expensive and wasteful of resources, and most importantly does not truly serve the health needs of our patients, the public. Thank you.

DR. LADENSON: Thank you. The final presentation will be by one of our co-chairs, Dr. Orloff, whom again I want to thank for his cooperation in orchestrating this symposium. And he's going to

summarize the FDA's perspective on the issues we've been talking about.

DR. ORLOFF: Thank you very much. Let me begin by thanking Dr. Ladenson and his colleagues for their participation today. I want to thank the FDA speakers for their clear and concise explanations of the agency's science-based standards for determination of therapeutic equivalence of drug products, including levothyroxine sodium drug products. And let me thank Rose Cunningham for her diligence and skill in actually bringing this meeting together.

Backing up a little bit, I want to begin by making clear that going back to our original 1997 action against the unapproved levothyroxine sodium drug products, the FDA acknowledged in several places in that Federal Register notice the importance of accuracy in dosing of levothyroxine for all of its indicated uses. That is to say we fully recognize then, as we do now, as I said a few moments ago, the importance of precision in dosing with levothyroxine. Always in the interests of patients, both young and old.

That Federal Register notice, as you know, cited multiple problems attributed to the quality of existing marketed products, including the market

leader. These included adverse events upon prescription refill with the same brand, and after switching brands. And these, if you will, spontaneous reports that in isolation would not necessarily have been an indication of problems with product quality were bolstered, or essentially affirmed in their validity, or in indicating that, by instances of formulation changes documented to lead to super-potency, and multiple instances of low potency and stability failures prior to expiry, necessitating millions and millions of pills being recalled. And so as a result of this, as a result of this hard evidence of problems with the quality of this class of drugs, we took the action to require NDAs in order to assure the purity, potency, and stability of these products.

So what the FDA -- this harkens back to Dr. Malinowski's talk -- what the FDA didn't know, and couldn't count on in the past, and therefore we as physicians didn't know and couldn't count on in the past with regard to these products included aspects of potency, specifically today, by that I mean at release, or when the patient went to pick up the product from the pharmacy; tomorrow, when the patient took the second dose, or the next day when he or she took the third dose, because we had no controls over

content uniformity; next week, because likewise we had no handle on the actual decay profiles of these products, or indeed their stability overall; and next prescription because we had no controls over lot to lot consistency. Likewise, we didn't know enough about the dissolvability, and thus the bioavailability or the availability of the content levothyroxine in these products.

I should note just here, going back to some of the things that have been said today, that we all need to be aware that older studies conducted assessing the effects of changes in dose, for example Carr's study, assessing equivalence, for example Mayor's study, were conducted with these products. And to my knowledge, in none of these studies as far as I understand was assay, was quantitative assay of the content levothyroxine in the products at beginning and end ascertained. I could be wrong. I see Dr. Sherman looking in his book. But I think that that's something that we must be aware of as we look back at our historical data. Not in any way to disagree with the position, again, that precision in dosing, consistency in dosing is of critical importance for the health of our patients.

I might also add, just again because I

don't believe patients should be overly alarmed by the concerns of their physicians, that unlike Digoxin, unlike warfarin, while precision in dosing over the long haul is important for levothyroxine, there in fact is no more ideal drug, if you will, for permissible variation around some stable mean potency because of the long half-life, and because a single dose to one side or another of the desired dose in fact doesn't hurt the patient.

So today we have manufacturing standards for our approved levothyroxine products. As you've heard multiply, these include potency standards whereby the historical overages that were put into the products to compensate for initial rapid levothyroxine degradation, are not permitted under the NDAs. The approved products, that is, must target 100 percent of labeled potency at release. Lot to lot consistency is controlled, and there are specifications on dose-content uniformity, that is to say the distribution of potencies around the mean. And again to repeat, in this day and age the mean for the product content within the bottle of levothyroxine that you get conforms at release within a couple of percentage points to what it actually says on the label. We never had that before. We have stability standards

such that the new products are limited under obviously the controlled conditions in which they're tested are limited to less than or equal to 10 percent loss of potency to expiry. That is to say, if appropriately cared for, they are labeled to contain up through their shelf life at least 90 percent of their labeled content. It is notable that because of overages, certain of the old levothyroxine products could lose as much as 15 to 20 percent of potency over their shelf life. So at this point, FDA is confident that any small differences in potency at release between levothyroxine products are not clinically important. Additionally, we believe that levothyroxine product potency standards at release and expiration ensures that products will remain safe and effective throughout their shelf life.

Well, what about the biopharmaceutical characteristics of these approved products about which we've been talking a lot today? Well, as Dr. Davit has explained and others, none contains excipients that were suspected to or have subsequently been shown to affect the absorption of the active ingredient. All of these products rapidly and readily dissolve in vitro and are presumed to do so in vivo. And, as has been stated a number of times, all of these

levothyroxine sodium tablet products approved to date essentially perform like solutions. That is to say, the levothyroxine content of these tablets is similarly bioavailable, absorbable by the patient, as it is in a solution of levothyroxine. And since all solutions of levothyroxine are by definition identical, then a priori we do assume that these products will indeed perform very similarly.

Notwithstanding that assumption, however, as you also know, we do require something called bioequivalence testing. And bioequivalence testing is applied both in the determination of therapeutic equivalence between drug products, and in the determination of dose proportionality within a drug product. And as I said earlier from the desk there, dose proportionality is something that's essential to our ability as thyroid physicians to accomplish the precision in dose adjustment on which we rely to titrate our patients to the thyroid hormone status appropriate to the condition being treated, and against symptoms and signs and laboratory signs of either hypo- or hyperthyroidism. In other words, in order for these products to be therapeutically useful, we require that evidence be presented to establish that when we increase the dose of levothyroxine, for

example, from 100 to 112 micrograms, 12 percent more levothyroxine is indeed bioavailable on average every day of therapy. In both cases, that is for the determination of therapeutic equivalence of two different levothyroxine products, and for the determination of the dose proportional bioavailability of two dosage strengths of the same product, the bioequivalence test is a confirmatory in vivo assay of product performance. As we've said many times, it looks at the rate and extent of absorption of active ingredient. It is always conducted on pharmaceutical equivalence. It is not conducted on two products that aren't pharmaceutically equivalent, and it follows a conclusion, and is considered in the context of that conclusion that dissolution characteristics and, parenthetically, differences in the excipient content of the products don't suggest a likely effect of formulation differences. And I should say, again, that these studies by their design, that is to say their sample sizes, by their analysis and interpretation fully recognize the impact of inter- and intra-subject variability on the absorption of drugs.

Well, the results of the bioequivalence tests that FDA has reviewed across different approved

levothyroxine drug products have shown that the observed differences between products we have deemed therapeutically equivalent in the rate and extent of absorption of levothyroxine, and the differences within products, where we've concluded dose proportionality across the approved dosage range, are of similar magnitudes and variability from study to study, and from drug to drug. And in all cases, these differences and the statistical 90 percent confidence intervals around them have all been well within FDA's limits of acceptance for clinical sameness, including for narrow therapeutic index drugs.

So we conclude from the bioequivalence data that we have reviewed that if there are any small differences in the performance between different dosage strengths of individual products, these differences are not clinically important, and you and I and our patients should feel confident that when we titrate the dose of levothyroxine, we are actually titrating the dose as it says on the label. We are further confident that if there are any similarly small differences in performance between products listed as equivalent, these are likewise not clinically important.

Let me step back for just a second for a

little bit more perspective. I think it is agreed around the room that the historical pre-NDA levothyroxine products were poor tools for the management of thyroid balance. I think we all understand that the quality problems associated -- or that characterized those products made them really less than ideal as therapeutic products for the treatment of our patients. And yet, notwithstanding the repeated problems in potency and stability in evidence based on analyses of the products and based also on problems that were faced by patients, all of which prompted our 1997 action, we were still successful overall in the treatment of our patients. Today, because of requirements imposed by FDA, the NDA approved and the ANDA generic approved levothyroxine products are far more reliable than the historical unapproved products. They are, number one, consistent across products in potency at release, and consistent across products in permissible loss of potency to expiry, although it is perhaps important for physicians to understand that some of the products lose potency faster than others.

This slide actually shows the expiration dates based upon stability testing. We've got one product that actually variably across the dosage

strength expires at nine months, and some of the higher dosage strengths have 14 months shelf lives. We have one product that expires across the dosage range at 12 months. We have three relatively more stable products with shelf lives of 18 months, and we have three of the most stable products with shelf lives of 24 months.

Finally, FDA has felt all along that the societies' concerns regarding the efficacy and safety of levothyroxine drug products that we have approved and deemed therapeutically equivalent arise because of a misunderstanding of the scientific basis for our determinations. The societies have also raised significant concerns among physicians and patients in this clinical area, which at least with regard to our therapeutic equivalence determinations, this has -- I'm not making any comments about switches for products that we have not deemed therapeutically equivalent, we do not believe are justified. And therefore, we think they're unfortunate.

It's been the goal of FDA's presentations here today to explain once again our methods and our standards. And I hope we've been clear. I also feel that we need to point out the absence of scientific evidence of risk or harm arising from these approvals,

and the therapeutic equivalence designations.

And so I must go back to the societies' position statement. First, the societies have asserted in their position statement risk of switching from old to new at the time of approval of the NDAs for levothyroxine, suggesting that FDA mismanaged that period of transition. But no evidence of risk or harm has emerged. Second, the societies have also asserted or concluded risk of switching from one product to its generic or AB rated equivalent where no scientific evidence of risk or harm has emerged. I think we need all to be clear here, notwithstanding Dr. Wartofsky's questionnaire presentation. The fact that pharmacists substitute is not evidence of risk. The fact that patients may not know it is not evidence of risk. The fact that patients may not have had their TSH checked in temporal relation to such a switch is not evidence of risk. And finally, anecdotes of change in thyroid status after a switch are likewise not scientific evidence of risk, i.e., directly implicating the switch in the change in thyroid status. Suffice it to say, and that's been part of the discussion here, and that's got to be part of the follow-up to this meeting, no formal studies of differences in efficacy, if you will, within versus across products have been

conducted, adequate and well controlled studies, although we welcome the societies to work with us to conduct well controlled studies to affirm our methods and designations. Although as I said before, this is not likely to come from the regulated industry, and I don't believe that FDA is going to be able to conduct those studies itself.

So in conclusion, FDA is confident of its methods, including its bioequivalence standards for determining therapeutic equivalence. Physicians and patients should likewise have full confidence in the quality of the approved products, and of the therapeutic equivalence of products so listed. FDA does not believe that any small differences related to potency or performance that may exist between products, within products across doses, or with aging, assuming appropriate care of the products by the patients, are clinically important, although we do believe it is important for physicians to understand that some products have shorter shelf lives than others, and thus some lose potency more quickly than others.

Finally, the risks as the societies construe them of alterations in thyroid balance associated with switching levothyroxine brands based

on FDA's designations are no different, we contend, than the risks, if you will, of refilling a prescription of the same brand of levothyroxine. I thank you for your attention. I gather we're going to break for a few minutes before we return for our final period of discussion. Thank you very much.

DR. LADENSON: We will break until 4:05 and then return for what Dr. Orloff and I -- will be a final forward-looking period of discussion.

(Whereupon, the foregoing matter went off the record at 3:50 p.m. and went back on the record at 4:52 p.m.).

DR. ORLOFF: Okay. Welcome back everybody. We're going to take this into the end of the day. I have a couple of people who signed up to speak in this session. The first is Dr. Robert Jerussi. Do you have comments you want to make, Dr. Jerussi?

DR. JERUSSI: I do.

DR. ORLOFF: Okay. That's fine. And Bill Landschulz is second. Three minutes, please.

DR. JERUSSI: Good afternoon. I'm a chemist and a consultant. I am being paid to be here. I have a client who's interested in this. But on a more personal note, I would say I would congratulate

all the physicians who are here for your dedication, and for your care for your patients. I'm really impressed with that today.

However I'd like to point out that in the 1990s, there were multiple recalls of lots of these products, dozens, by the dozens they were recalled. And some of the companies were in the position where they had decent stability, better than some of the results you saw on the slide here, and had validated the manufacturing process, and a year later things went like this, with no explanation. That hasn't been completely explained. So FDA did a lot of monitoring at that time, and the question I have for FDA, are you monitoring today what you've recently approved, especially those with short-term batches? Secondly, how many recalls have you had? I haven't looked that up. The old recalls are on the internet. How many recalls have you had of the presently approved material? I think those numbers are important.

And secondly, as to the affected patients, Dr. Orloff said somehow you managed to take care of your patients during the 1990s when things were sort of haywire. What is the average adverse reactions in the `90s compared to from 2000 on?

There should be some idea whether all this work really improved things for patients.

DR. DUFFY: As far as recall data, I'm not familiar with the recall rates, and whether they're different than before. But that's something we can look into.

As far as monitoring the product quality in the marketplace, we are doing that. We have a standard program in place to assess the quality of product we get right off the shelf. And we have been monitoring that. And they have been shown to be suitable quality in the marketplace. We have those data.

DR. ORLOFF: That's for all products, right Eric?

DR. DUFFY: That's correct.

DR. ORLOFF: This is not just uniquely for levothyroxine.

DR. DUFFY: Not unique to levo. We have a not quite random -- we select products based upon potential for problems that we might be aware of, and levothyroxine is one that we wanted to see whether these changes had in fact resulted in a better quality product. And it appears that that is the case.

DR. LADENSON: I'd like to just respond to Dr. Jerussi's question about adverse events if I might. And that is to suggest that it would be very hard on an anecdotal basis to know whether there were or are more adverse reactions. The kinds of reactions that we're talking about are non-specific symptoms, common clinical events like atrial fibrillation and myocardial infarction that have many different etiologies. And I think in the same way that it might be hard to see the level of the ocean rising a millimeter or two, it would be hard to know how levothyroxine therapy was contributing to those. I think one only needs to see the recent experience with the COX-2 inhibitors, for example, to see that that was not something that came to light by virtue of a broad societal or medical recognition of the complication, but rather only with rigorously controlled observations. I don't know whether, David, you have any thoughts about that.

DR. ORLOFF: Bill Landschulz. And Sally Schimelpfenig is next.

DR. LANDSCHULZ: Hi, I'm Bill Landschulz. I'm from Abbott Laboratories, the Clinical Development group. There has been some conversation about dissolution and other in vitro assays. I'd

like to just point out that dissolution per se does not necessarily predict bioavailability, and that Synthroid has a very well characterized bioavailability. I think some of the conversation that we had with regard to the solubility of levothyroxine and counter anions, the pH and how it affects that can interfere with the assessment of bioavailability.

Of course it's the task -- as we have a very well characterized bioavailability, it is the task of the AB applicant to match that reference bioavailability, and to use Dr. Collins' comment that it is not statistically significantly different in bioavailability. Presumably, statistically significant means clinically significant as well, and I would argue that clinical significance is most likely visualized by evidence of risk. Now, we appreciate that finding the evidence of risk is going to be difficult, just to Dr. Ladenson's comment that it will be very difficult to see changes in adverse events in things that are either very subtle, like children's IQ, or very prevalent, like heart disease.

We appreciate that Dr. Orloff points out that the width of the goalposts can easily be subverted by simply increasing the size of the number

of subjects in the study. So perhaps we should be thinking about it a little bit differently, and picking on a comment that you made about precision of dosing, from refill to refill, I'd agree that that probably is the key question. So let's put aside what the marker would be. We can decide whatever that marker is. But I think the real question then would be what is the necessary precision of dosing that we need to meet from refill to refill? Is it 9 percent? Is it 10 percent? Is it 12 percent? Is it more than that? I think that's an important question that I hope that we all can come to consensus on soon.

DR. ORLOFF: Looks like we have another speaker.

MR. POMERANTZ: Good afternoon.

DR. ORLOFF: Please state your name.

MR. POMERANTZ: I'm not Sally Schimelpfenig.

DR. ORLOFF: No, you don't look it.

MR. POMERANTZ: My name is Eric Pomerantz, and I'm with Sandoz. I would just like to take an opportunity to thank the members of this panel, and the members of the FDA today, to allow us the opportunity to present our collective knowledge

and experience developed in pursuing an AB rating for our NDA-approved levothyroxine product. Thank you.

We commend the FDA and its dedicated scientists and clinicians for their devotion to public health priorities in levothyroxine and all other regulated products. I think a consensus has emerged today, that any product, whether the brand an AB rated brand, or an AB rated generic ANDA can provide patient benefits if used carefully and monitored properly by physicians. Sandoz looks forward to continuing to work with the FDA in a meaningful way as we pursue our goals of serving our patients by enhancing patient access to competitive products. Thank you again. I appreciate that we were able to come, and I think I speak on behalf of the others in industry that we were given the opportunity to participate today.

DR. ORLOFF: Thank you very much. Dr. Ladenson, would you like to get us started on the? We're going to try to open our final discussion here.

DR. LADENSON: What the societies wanted to suggest for our home stretch discussion was to return to the goals that we came to the meeting with, and discuss the feasibility of addressing them

together. And I would remind you what those were: to look at the feasibility of making more stringent the bioequivalence standards or goalposts; to assess the value of adding TSH as a pharmacodynamic measure, and perhaps testing the hypothesis that some have questioned today of its value in assessing the therapeutic equivalence of thyroxine preparations; to hear a bit more from the FDA about what regulatory powers it has, if any, to strengthen adherence to laws regulating switching by non-prescribers; and then finally, I think to really devote a little bit of time to talking about the feasibility of designing a definitive trial with appropriate controls to test some of these hypotheses, that narrower goalposts are required and appropriate, that TSH would be a welcome addition to equivalence assessments.

And so I guess maybe an easy one to address that I'd be interested in hearing from FDA about are what its powers are with regard to warnings and regulation of switching behavior.

DR. ORLOFF: I'm not going to call any of FDA's attorneys up to the table here. I think what some of us were talking about before this final session is that we believe that, at least it sounds as though there is significant confusion out there as

to what products are indeed rated AB equivalent, and what products are not yet rated AB equivalent. We talked earlier about the complexity of that matrix, and about I think in the short run at least the poor feasibility of expecting that it would be completed in a formal sense. So I think what we can do, the FDA back at our place, is to work to develop perhaps on our website some clearer information and delineation of exactly what products are AB rated one to the next, much as the societies have included in their position statement which issued at the end of last year. But I think that we can play a role in disseminating that information better, perhaps, or making it more readily available so that if indeed some of this confusion, or some of this switching is at least according to our designations inappropriate, that we can stop that. But I don't believe we can go out and enforce -- we don't have an enforcement function on the practice of pharmacy in that sense, the dispensing of drugs.

DR. LADENSON: Dr. Conner?

DR. CONNER: I can speak not so much as an FDA person but as a pharmacist that a lot of the concerns that we've heard mainly are, as Dr. Orloff said, the practice of pharmacy, which is regulated by

the states. And that's why you saw from the slides and the discussion of various state rules. Each state has some different rules as far as what prescribers are able to pre-specify, and what pharmacists are allowed to switch to or from. And you know, the FDA doesn't have any direct power over that. But of course, as always, we have an educational role, and an educational responsibility, and we can certainly influence the switching and prescribing in that way. But as far as direct regulation of how pharmacists switch, or perhaps the major motivating factor behind pharmacists switching which is what various payment plans either allow or mandate as far as what the patients are allowed to get, which is perhaps an even more compelling reason than pharmacists and pharmacies wanting to make a profit. I think that's -- overall the more compelling issue is the large payment plans and what their rules are.

DR. LADENSON: So that FDA would be in a position to more widely disseminate the relationships and how they exist. And would that be solely on a website, or is it something that you could discuss internally in terms of some kind of advisory to pharmacies? Do you ever issue such advisories?

DR. CONNER: Well, as far as -- this is speaking only for the Office of Generic Drugs. We have an educational program which we've gotten funding from Congress for to educate the public and physicians and other health professionals about generic drugs and what the standards are, and in part to give them a better feeling of confidence about the generic drug program overall by increasing understanding. So we have been given separate money to do those type of programs in the past. I don't know about for this specific question what would be possible or not, but it has been done.

DR. ORLOFF: I think we can simply commit to go investigate what our capacities are, and obviously we'll do what we're able to, and to the extent that we think it's appropriate we'll confer back with you.

DR. LADENSON: Dr. Hennessey?

DR. HENNESSEY: I just want to make a comment exactly to that. It is an extraordinarily confusing situation. If you simply look at the AB2 rated drugs, you'll find that, yes, each of the three major generics are AB2 rated, but for example, the Mylan product and the Sandoz distributed product are BX to one another. And the Unithroid is BX to the

Sandoz. And to think that a pharmaceutical distributor will substitute any of those three for an order for, let's say, Synthroid, but indeed would not necessarily in the next go-around respect the BX part is what I would assume would be the situation. I think it's an extraordinarily confusing situation.

DR. CONNER: Well, this is purely guesswork on my part because I wasn't around when the whole system of organizing the AB ratings, and listing them, and how the Orange Book was organized, but it seems to me that the whole system was designed with a more simple situation in mind. I mean, you have one reference-listed drug that's approved through an NDA process on which clinical trials, and you have a number of AB rated generic products that are properly approved based on that original product. I mean, for that type of situation which is most of the things we do, the system works very well, I think.

We have a number of products, fortunately it's not a huge number, where it becomes a bit more confusing, where you have several NDAs for the same drug substance, but they have different labeling, perhaps different indications, and so forth, and so we've had to go to this AB1, AB2, and so forth to

distinguish officially between generics that only should be substituted for that one. So levothyroxine isn't the only one, but it isn't a huge list. And it's trying to make a system that may not have been designed for that work with a much more complex situation. And so obviously the more complexity you put into the system, the more confusing it gets for people who just barely understand it.

DR. HENNESSEY: And that's exactly what one of our concerns is, is the complete confusion in the marketplace where every time the patient walks in they may walk out with a different shaped pill, generating more phone calls, etcetera, etcetera. And when we look at the spectrum of differences among the AB2's for example, ranging from 12.5 percent difference in bioavailability down to around 3 percent difference in bioavailability, there may be differences amongst the generic substitutables. So I don't know.

DR. CONNER: Well, I mean that's the -- different appearance of different products, brand name and generic, I mean is a problem -- I wouldn't say it's a problem. It's a characteristic across the board. I mean, every manufacturer -- and it's a good thing, because every manufacturer has their own

market image, their own type of tablet, and that way you can actually look at the tablet and trace it back to who made it, and what strength it is, and so forth. So it actually is a good thing. However, I think anytime you go into your pharmacy and you come out with a different color tablet, or a different shaped tablet, some patients that haven't been assured that yes, this is the proper generic, you've been given the proper strength and so forth by the pharmacist, you know, has questions. So that is a characteristic, or a question of patience. And doesn't really even put it -- you know, it's not putting into question whether they're really getting an equivalent product or not, but I have -- I've just gotten something different, and I have some doubts.

DR. HENNESSEY: Generating a lot of confusion.

DR. CONNER: Yes.

DR. LADENSON: I'd like to --

DR. ORLOFF: Before we go on, I just want to say, so the resolution of this question is that we'll go back and look into it, but the society should understand that our position stands; that we believe that those products that we've rated as AB equivalent are indeed AB equivalent, and we're not

going to issue any kind of public education or whatever that says don't accept a substitute of those things we've designated as therapeutically equivalent. So I know that's not going to satisfy you fully, but we can address some of the complexity by trying to make clear which ones have officially been designated as equivalent.

DR. LADENSON: And I'd like to, on behalf of the societies, suggest that we will certainly be interested in cooperating with you in that. And I think one can envision a site that would be accessible to patients, and perhaps linked to by all of our sites and the patient education sites that would allow people to ask questions. Is what is being proposed as a switch for my prescription, what category is that in, and what does it mean for me. So we'd be very interested in cooperating with you on that.

DR. ORLOFF: To some degree -- I don't want to get into details of it now, but to some degree, obviously, the reciprocity, or the linking of those two sites is going to require some agreement on the fundamentals here. I'm not sure we're going to come there. That's not to say that having, you know, a link from your site to our site is not

inappropriate, but I'm not positive we can do it the other way.

DR. LADENSON: Right. And it might even include the ability to identify tablets so that patients would be able to know that they were on A and were being switched to B, and then find out what that meant in terms of your advice.

The second issue I wanted to ask FDA about was what it would take to narrow the goalposts. Does this require a large study, or is it not possible, given the concerns of clinicians, and your own previous statements about what you consider appropriate for this narrow therapeutic index drug, to simply decide that 80 to 125 percent is too broad for this drug. What are the obstacles to that?

DR. ORLOFF: Well, again, we're going around and around here. By and large, with one exception that you've seen, the 90 percent confidence intervals around the means for the ratios of the AUC zero to 48's and from the levothyroxine bioequivalence studies across products already fall well within the 80 to 125 tolerance limits. So I'm not exactly sure what narrowing the goalposts is going to mean. As I said before, and I think it's absolutely true, if we want to narrow the confidence

limits, or if -- let's just say if anybody wants a narrower looking 90 percent confidence interval around the mean, all we need to do is do larger studies. So I'm not sure that that is really not the solution here. The societies, I believe, are focused on the mean, the point estimates for the differences in these single studies, in fixed number of patients, where there is no adjustment for baseline potency, and where, as I said, there are a lot of priors going into it, like pharmaceutical -- by and large pharmaceutical equivalence and dissolution.

So I don't think -- I guess I would say that we shouldn't go to the question of narrowing the goalposts, because I don't think that's the solution here. I actually think, if I might, Dr. Ladenson, that we ought to spend the time talking about what would be the aspects to brainstorm here -- what would be the aspects and the practicalities behind doing the confirmatory study, or as I said before in the made-up word, the refutatory study, to examine the integrity of our determinations, or the legitimacy of our determinations from a clinical standpoint. And I believe that that study can only be done at, and you believe too, that it has to be done as a TSH study.

Now we would not be conceding, in working

with you on such a study, that we do not adhere to what we've said all the time here. We would not change our regulatory position, or our regulatory procedures in the meantime. But we do believe that we are at an impasse here, sort of at an intellectual level if nothing else, and it needs to be resolved. And the only way to resolve it is to work together to get the right study done.

DR. LADENSON: Before we put the goalposts aside, I'd just like to point out that one thing that FDA could do that would be very reassuring to the clinical community would be to say `We see why you're uncomfortable with a drug that is the most commonly substituted drug for a currently prescribed drug. We understand why with the 90 percent confidence limits being 22 percent, you and your patients are worried, and we see an opportunity to make a modest change that would at the outset be really pretty reassuring to patients and physicians.' And now I'm happy to put it aside.

DR. ORLOFF: Okay. Well, fine. Let's move on.

DR. LADENSON: And I hope you'll think about that. The big point, as David -- yes, Dr. Ridgway.

DR. RIDGWAY: Well, to go to David's point, I think one of the things that the societies are looking back at you with is what are going to be the ground rules here for this study. I mean, it's very interesting to look at the societies and say, okay, let's perform this study. You guys perform it and pay for it, but what are going to be the ground rules for change if it's refutatory? If you do a steady-state study, what are going to be the ground rules for what is significantly different? And I'd like to talk about that, to see what that would be. Because there's no sense doing a study if whatever we come up with is not going to be deemed as valid.

DR. ORLOFF: Well, actually I don't think that that's a fruitful approach to this. I think that we need to agree to work together to design a scientifically valid unbiased investigation to the best of our ability. We cannot commit here to contributing funds to the conduct of such a study --

DR. RIDGWAY: I didn't ask for funds, David.

DR. ORLOFF: Okay.

DR. RIDGWAY: I didn't ask at all for funds.

DR. ORLOFF: Furthermore, Chip, we cannot

commit on the basis of whatever hypothetical result the study shows to some change. Let's just say from where we are concerned, speaking for those of us around the table and for the agency, were an unbiased scientifically valid study to definitively refute our methods, we would all be in shock. That's where we stand. So we are very interested in working with you, but I think it's far too much to ask that we could now lay out a series of, you know, a decision tree based upon what the hypothetical results might be. So I think we need to first begin by looking at what the design of such a study would be, what the hypothesis testing potential, or what the hypotheses are we want to test, and how to design a study to test those hypotheses. And then, move from there to the conduct of such a study. The results will be what the results will be. And we'll look at them and take them under consideration, all of us.

DR. RIDGWAY: Okay, David, that's fine. But what you're basically saying is that the FDA would be in total disbelief if such a study showed that your current procedures were refuted. If I am quoting you correctly.

DR. ORLOFF: That is our --

DR. RIDGWAY: That's the hypothesis we're

testing?

DR. ORLOFF: No, that is not the hypothesis we're testing. Please, don't take my words and turn them around. What I said is we cannot commit to -- we can't have a discussion about what we would do as a result of such a study not knowing what the results of the study are. Okay? How about this. Should the results of a valid study refute our methods, then clearly we would have to reevaluate our methods. Should the results of such a study confirm our methods, then clearly the societies would have to reexamine their understanding, and their interpretation of our AB ratings. So it goes both ways. That's what we're trying to work together.

DR. RIDGWAY: Unequivocally, and I think every society speaker has made that point. That second point that you just made.

DR. ORLOFF: So, but the only path forward here is to work on designing the study and getting it done.

DR. LADENSON: Steve?

DR. SHERMAN: One of the parts behind Chip's question might be the statistical one, which is without having a sense of what magnitude of difference is going to be viewed as relevant to the

discussion, it's hard to power a study to minimize the beta error. So one has to work towards some agreement as to what would be a relevant difference to be looking for.

DR. ORLOFF: Well, that is obviously a critical detail of such a study. I don't know that we're going to resolve that specific detail here today. I wouldn't even propose to get into it. I think that probably the best we're going to get into today is to resolve to convene some sort of working group to move ahead to try to develop the study to examine the issues that need to be considered in this hypothesis test.

DR. WARTOFSKY: We would be delighted to join in a working group to pursue this, but I think one of the basic issues here is we continue to be talking apples and oranges, different things. What is the definition of the FDA methods assessing bioequivalence? You said you would be shocked or surprised if anything was refuted. Depends on the definition. You -- in your talk, you concluded that there was clinical sameness. There isn't clinical sameness. There's pharmaceutical sameness. On the basis of the bioequivalence data, you don't have the authority to say that there's clinical sameness, or

that there is no difference in clinical outcome. We're seeing the clinical outcome. There is a difference in clinical outcome. So it would depend on the definitions, and how the study is done, what we're looking at. I wouldn't be surprised if the bioequivalent data is exactly confirmed. But the issue is what is the therapeutic equivalence. That's where we're having a disconnect.

DR. ORLOFF: No Len, we actually -- we're talking here about committing to work towards a TSH based study. But I do -- I think you need to be very careful with your words about authority, and about our scientific conclusions. You do not have evidence of risk. You have anecdotes, and you have a wholly unscientific data-gathering process whereby you've biased beforehand your societies by issuance of a position paper, and then asked them whether they're concerned about the issue. A 5-page position paper in which you tell them over and over again how incredibly dangerous this problem is, and then asked them whether they think it's dangerous. That is not a study. So I think you need to be very, very careful.

There's a tremendous amount of alarm here, and what we're talking about, and that's where

we need to come -- we are going to have to agree to disagree at this point. And we're going to have to send you and me back, and every other doctor in this room, to manage our patients the way you've been managing them yesterday and the day before. And if that involves some phone calls of concern, either legitimate or non-legitimate, depending upon where you stand, we're just going to have to deal with that. But in the meantime, as I've said before, the only path forward here is to figure out how to do a study to ask the question as to whether these things are clinically identical. Okay? That's your question. And we, of course, take the position that our standards define clinical sameness, but you don't agree with that. We understand. Okay? So we now have to -- and we also understand that as practitioners we follow our patients with TSH levels. And we understand that that is, for the purpose of using the drugs, that is the clinical endpoint of interest, and it is in truth the only way to definitively establish whether our methods hold up, or whether they don't hold up. So I guess we're going to just have to agree to work together to convene something. I don't know that we're going to be able to nail down any specific issues today, but

go ahead.

DR. WARTOFSKY: We do agree to disagree, but when you say in one of your last slides that there is no risk proven to switching. There is no risk proven to not measuring a TSH. There is no risk proven to not re-titrating, whatever. If you cross Independence Avenue against a red light, you get hit by a car. Observable. If I cross, is there a risk to me? I'd say the red light is analogous to the TSH. We see a TSH go from 1 to 9 with a switch, crossing the red light. We see a TSH go from 1 to 9 in a pregnant woman, and she delivers a fetus at risk. It's logic. Some things you just cannot prove without doing the large studies that we don't have the data.

DR. LADENSON: I think one important part of such a planning group would be to what degree to accept TSH as a surrogate for rare adverse events. Is one way to perhaps put what you're saying. And I think that would require extended discussion.

DR. ORLOFF: That's the question of what the goalpost is for a difference in TSH at the end of the day. And that's something we'd have to discuss. What is a clinically significant difference in TSH. How much would you be willing to accept every six

months as a variation in a given patient as not meriting re-titration of their drug.

DR. LADENSON: David, are there precedents for what's being proposed here, where FDA has collaborated not in terms of defining a trial that industry had to carry out, but that a professional society was to pursue to test hypotheses about the adequacy of current let's say regulatory standards?

DR. ORLOFF: I am not aware that there are precedents. I think -- I'm not sure that it matters whether there are precedents. What matters is that we do a scientifically valid study. Or we work together towards the completion of a scientifically valid study.

DR. LADENSON: Dr. Ridgway?

DR. RIDGWAY: Just one point. We at the table have actually talked about this TSH variability a lot. And we actually have some ideas about what would be the goalposts. But I do want to remind the audience today, and certainly the people at this end of the table that what we've tried to present today is not biased stuff. This is not data that I generated, or a drug company generated. This is data that is in the literature about risk being associated

with toxicity. And when we get that list, David, we have not produced any evidence of risk with these statements, FDA likewise has not proved one bit of evidence of safety by their standards in this area.

DR. LADENSON: The format of such a working group, how would you picture that working, David, at the initial phase?

DR. ORLOFF: Well, I gather -- I think that in any of these collaborations that go on across the great USA we're lucky we have email, and faxes, and phones. And I'd propose that we probably begin by a brainstorming exercise, that we're not going to conduct today, but whereby we sort of throw our ideas into the ring as to what factors need to be taken into consideration in study design. And I think at that point we need to go from there.

With regard to the logistics of the actual conduct of such a study, as I've said, we can't, sitting here today commit to anything, although that's not to say that we cannot investigate FDA or some other aspect of HHS's contributions to such an investigation.

DR. LADENSON: Are there other comments from the panelists or the audience? Well, I'm glad that we are ending on what I consider, at least, a

positive note. And I'm sure that the societies are going to want to pursue this. And you can expect to hear from us within a fortnight.

I also want to just say that I'm impressed, and I hope the other speakers and the audience are impressed by the sincerity with which everyone who has been a part of this meeting has approached the issues here. And I think all of us share a common concern for the Americans and others in the world who take thyroxine. And I think if we stick with that in mind, we could make this collaboration a profitable one.

DR. ORLOFF: Let me add my thanks to all those who participated. I do believe it was fruitful, if not contentious. And we will have to agree to disagree on some of the issues. I guess from this point on I encourage rigorous, hard science across both sides of this. And we will hope that in time we can accomplish what we've set as our goals. Thank you everybody.

DR. LADENSON: I want to especially thank Rose Cunningham and Bobbi Smith and her team for putting together the meeting. Thank you.

(Applause)

(Whereupon, the foregoing matter was

concluded at 4:52 p.m.).