DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

JOINT SESSION WITH THE

NONPRESCRIPTION AND DERMATOLOGIC DRUGS

ADVISORY COMMITTEE

VOLUME I

Wednesday, March 23, 2004

8:00 a.m.

Hilton Washington DC North
620 Perry Parkway
Gaithersburg, Maryland

P A R T I C I P A N T S

Alastair Wood, M.D., Chair

Shalini Jain, PA-C, Executive Secretary

Committee Members:

Michael C. Alfano, DMD, Ph.D., Industry Representative
Terrence F. Blaschke, M.D.
Ernest B. Clyburn, M.D.
Frank F. Davidoff, M.D.
Jack E. Fincham, Ph.D.
Sonia Patten, Ph.D.., Consumer Representative
Wayne R. Snodgrass, M.D., Ph.D.
Robert E. Taylor, M.D., Ph.D., F.A.C.P., F.C.P
Mary E. Tinetti, M.D.

Special Government Employee (Voting):

Michele L. Pearson, M.D.

Government Employee Consultants (Voting):

John S. Bradley, M.D.
John M. Boyce, M.D.
Ralph B. D'Agostino, Ph.D.
Thomas R. Fleming, Ph.D.
Elaine L. Larson, R.N., Ph.D.
James E. Leggett, Jr., M.D.
Jan E. Patterson, M.D.

FDA Participants:

Tia Frazier, R.N., M.S.
Charles Ganley, M.D.
Michelle Jackson, Ph.D.
Susan Johnson, Pharm.D., Ph.D.
John Powers, M.D.
Curtis Rosebraugh, M.D.
Debbie Lumpkins, Team Leader

C O N T E N T S

P R O C E E D I N G S

Call to Order and Introductions

DR. WOOD: Let's get started. Welcome to the Over-the-Counter Advisory Committee. Let's begin by going around the table and everybody introducing themselves, and we will start on this side, Charlie.

DR. GANLEY: Charley Ganley, Director of OTC.

DR. POWERS: John Powers, Lead Medical Officer for Antimicrobial Drug Development and Resistance Initiatives in the Office of Drug Evaluation IV.

DR. ROSEBRAUGH: Curt Rosebraugh, Deputy Director, OTC.

DR. JOHNSON: Sue Johnson, Associate Director, OTC.

DR. LUMPKINS: Debbie Lumpkins. I am a Team Leader in OTC.

DR. DAVIDOFF: I am Frank Davidoff. I am an internist and editor emeritus of Annals of Internal Medicine and a member of the OTC

committee.

DR. FLEMING: Thomas Fleming, Chair, Department of Biostatistics, University of Washington.

DR. FINCHAM: Jack Fincham, professor at the University of Georgia, College of Pharmacy, and I am a member of the committee.

DR. CLYBURN: I am Ben Clyburn. I am an internist at Medical University of South Carolina and a member of the committee.

DR. BRADLEY: I am John Bradley, a pediatric infectious disease doctor from Children's Hospital, San Diego, and I am a member of the Anti-Infective Drugs Advisory Committee.

DR. PATTERSON: Jan Patterson, Infectious Diseases and Infection Control, University of Texas Health Science Center, San Antonio and South Texas Veterans Healthcare System.

MS. JAIN: Shalini Jain, Acting Executive Secretary for today's meeting.

DR. PATTEN: Sonia Patten. I am the consumer representative on the panel, and I am an

anthropologist on faculty at Macalester College in St. Paul, Minnesota.

DR. SNODGRASS: Wayne Snodgrass, pediatrician and clinical pharmacologist at the University of Texas Medical Branch.

DR. LARSON: Elaine Larson, from the School of Nursing and School of Public Health at Columbia University, in New York.

DR. TAYLOR: Robert Taylor, Chairman, Department of Pharmacology, Howard University, in Washington, internist and clinical pharmacologist.

DR. BLASCHKE: Terry Blaschke, internist, clinical pharmacologist, Stanford, member of the committee.

DR. TINETTI: Mary Tinetti, internist, Yale University and member of the committee.

DR. D'AGOSTINO: Ralph, D'Agostino, biostatistician from Boston University, consultant to the committee.

DR. LEGGETT: Jim Leggett, infectious diseases at Portland Medical Center and Oregon Health Sciences University, and I am a member of

the Anti-Infective Drugs Advisory Committee.

DR. ALFANO: I am Mike Alfano, New York University College of Dentistry, industry liaison to NDAC.

DR. WOOD: And I am Alastair Wood and I am the Chairman of the NDAC and Associate Dean at Vanderbilt.

So, let's get started. Shalini, do you want to read the conflict of interest statement? While she is digging that up, the weather has caught us and the first speaker from CDC is stuck in Atlanta--the story of people's life in the Southeast. So, what she is going to do, she is on her way back to her office and she is going to e-mail us slides and then we will try and project the slides later in the morning, with her talking to us over the telephone. So, that will be a nightmare I suspect.

[Laughter]

That means we will time shift everything up and then probably, depending on how she gets on, we may have the question and answer period for the

first ones a little bit earlier and take an earlier break and then come back to hear her, depending on how the technology is behaving. Shalini, go ahead.

Conflict of Interest Statement

MS. JAIN: The Food and Drug Administration has prepared general matters waivers for the following special government employees who are attending today's meeting of the Nonprescription Drugs Advisory Committee on the microbiologic surrogate endpoints used to demonstrate the effectiveness of antiseptic products used in healthcare settings. The committee will also discuss related public health issues, trial design and statistical issues.

This meeting is held by the Center for Drugs Evaluation and Research. The following meeting participants have waivers: Dr. Jan Patterson, Dr. Sonia Patten, Dr. Thomas Fleming, Dr. John Boyce, Dr. Ralph D'Agostino and Dr. John Bradley.

Unlike issues before a committee in which a particular product is discussed, issues of

broader applicability such as the topic of today's meeting will involve many industrial sponsors and academic institutions. The committee members have been screened for their financial interests as they may apply to the general topic at hand. Because general topics impact so many institutions, it is not practical to recite all potential conflicts of interest as they apply to each member. FDA acknowledges that there may be potential conflicts of interest but, because of the general nature of the discussions before the committee, these potential conflicts are mitigated.

With respect to FDA's invited industry representative, we would like to disclose that Dr. Michael Alfano is participating in this meeting as a non-voting industry representative, acting on behalf of regulated industry. Dr. Alfano's role on this committee is to represent industry's interests in general and not any one particular company. Dr. Alfano is Dean, College of Dentistry, New York University.

In the event that discussions involve any

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

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

DR. WOOD: Thanks a lot. Let's go straight on to the first presentation from Susan Johnson. Susan?

Issue Overview

DR. JOHNSON: Good morning.

[Slide]

My name is Susan Johnson and I am the Associate Director of the Division of OTC Drug Products. On behalf of the division, I would like to welcome the members of the Nonprescription Advisory Committee and the Anti-Infective Advisory Committee and our other guests. As I am sure the committee members would agree, the bulk of the

background package as a metric of the challenge that we face today is certainly significant, and we certainly appreciate everyone making as much headway as they could with that background package.

We very much appreciate all of your assistance today. There is a wide variety of issues to discuss and so you will see the representation of the committee being broadened from NDAC to include the Anti-Infective committee members, and we appreciate everyone's attendance, as well as our consultants.

I will just be providing a brief introduction to the regulatory issues associated with the efficacy of OTC healthcare antiseptics.

[Slide]

The OTC healthcare antiseptics include three categories of drug products, the healthcare personnel handwashes; surgical hand scrubs; and patient preoperative skin preparations that are used to scrub the skin prior to surgery.

[Slide]

FDA's current approach to the evaluation

of healthcare antiseptic efficacy assumes that healthcare antiseptics play a critical role in infection control, and Dr. Michelle Pearson and Dr. John Boyce will discuss this role in additional detail. However, the efficacy of individual products must be demonstrated to meet regulatory requirements. FDA's current regulatory standards are based on actual product performance and have been supported in previous public discussions such as this one. Ms. Tia Frazier will explain more about the regulatory history of these products.

FDA currently determines the efficacy of healthcare antiseptics using a surrogate endpoint, and that is used as the reduction in a log

10 count

of bacteria from the site of the test product application. Dr. Michelle Jackson, from the Division of OTC, will discuss how the standard is used in the test methodology.

[Slide]

This meeting has been convened because we have received citizen petition requests to change the threshold criteria for bacterial reduction. We

wish to present our review for your consideration of the efficacy data in the literature for these products. We are asking that the advisory committee provide input about the standards that FDA needs to have in place to make regulatory decisions.

[Slide]

What are some of the factors that can influence efficacy of the healthcare antiseptics? This is by no means an exhaustive list but is intended to give you an idea of why product testing is required to demonstrate efficacy.

The first group of factors I am going to discuss are associated with the actual product. The active ingredient obviously affects efficacy. The spectrum of activity for each individual active ingredient is tested in associated testing criteria in vitro. The potency or dose response of the active ingredient shall also be taken into consideration, although in some cases it is not well known.

The formulation of the product can impact

its efficacy and influence that to increase or decrease efficacy so the concentration and dose delivered to the site and vehicle and other inactives in the products can affect efficacy. One thing that influences efficacy quite a bit is how the product is actually used, and that is led in large part by the way the product is labeled.

[Slide]

Other factors that influence efficacy of healthcare antiseptics include actual use parameters, adherence to the labeling and other practice standards and actual implementation of both labeling and practice standards.

There are many patient parameters that can affect the efficacy of these products, including things like health status which influences the risk for infection, as well as the type of procedure that is being conducted.

Resident and transient bacteria, resident bacteria being normal flora and transient bacteria being those sorts that are introduced during healthcare processes, can affect efficacy as well.

The amount of bacteria that is delivered and that resides on the skin, either prior to or that is left residually after product use, is an important determinant of overall efficacy. Virulence of the bacteria that exists on the skin affects efficacy as well. A small amount of bacteria can be present and provide a great risk of infection.

[Slide]

FDA in general assesses efficacy using randomized, controlled trials for the most part. These provide analytical strength and can be designed to control for multiple confounders. Critical to the design of controlled trials is the selection of active and vehicle control, and we will be discussing that later today.

[Slide]

The endpoints that are normally used in randomized, controlled trials are clinical or surrogate endpoints. Randomized, controlled trials typically use clinical endpoints because the relevance is more evident. In some situations the difficulty and expense of conducting clinical

trials is very important to industry. An alternative to clinical endpoints is surrogate endpoints, and Dr. John Powers will later discuss the scientific and regulatory precedent for using surrogates. Just as a reminder, and I am sure you have gleaned this from your reading already, but the current standards for OTC healthcare antiseptic efficacy are surrogate endpoints.

[Slide]

The factors that should be considered when using a surrogate to assess healthcare antiseptic efficacy include validity. We acknowledge from the outset of this discussion that there is limited information about the links between clinical outcomes and efficacy and use of the surrogates to determine efficacy. Dr. Steve Osborne will discuss the literature surrounding this question a little bit later.

The existing trials in the literature are not designed to validate our practice standards. Instead, our practice standards and use of surrogate are based on the use of antiseptics in

practice and our experience with marketed drug products.

Test methodology is also an important factor to consider when using surrogates. Test methodology should evaluate the conditions of use, largely directed by the labeling or the intended labeling. The test methodology to evaluate healthcare antiseptics with surrogates needs to characterize the tolerability of drug products. While we are talking primarily about efficacy today, the tolerability of these drug products is a major safety concern and does come up as part of the testing methodology. Test methods do need to be standardized with regard to all inherent procedures.

[Slide]

Other factors that should be considered when using surrogate endpoints are the decision thresholds and, as I have said, the current criteria are based on the NDA performance of existing approved products. We suggest that any changes to these criteria on decision thresholds

should be data driven.

Analysis of test data is critical, and later today Dr. Thamban Valappil will be discussing the analysis of these data. His talk is predicated on the previous discussions that we will be having about validity methods and thresholds, and he will talk about the need to evaluate the response of test products in the context of variability in both test methods and in patient response.

[Slide]

Epidemiologic studies do provide information for healthcare antiseptics. They provide actual use information on large populations and can often be used to suggest practice standards. They are often used to generate hypotheses to be later studied in randomized, controlled trials. But they are relatively insensitive to treatment differences and changes in things like threshold criteria. So, using them to extrapolate for regulatory decision-making is of limited value.

[Slide]

What specifically are we asking the advisory committee to address? First, can we continue to rely on surrogate markers to assess

healthcare antiseptic efficacy? I would like to remind the committee, as we will several times today I am certain, that we have the need for ongoing assessment and decision-making of these products so we do need to have standards in place now and in the near future, as well as into the distant future.

If surrogates can be applied, at least in the short term, is there compelling evidence to change our surrogate efficacy criteria now? What is the best way to analyze the efficacy data? And, what labeling information would be helpful for clinicians to understand product efficacy and potentially to compare among different products?

With that, I will turn it over to Tia Frazier, who is a regulatory project manager in the Division of OTC Drug Products, and she will be discussing regulatory history.

DR. WOOD: Just before you take that slide

off, there is sort of an underlying assumption there, which I think is right but I just wanted to articulate that there is a sort of regulatory inertia which is that in the absence of evidence we shouldn't change criteria. Is that fair? I am not disagreeing with that, I am just trying to put number two in that context.

DR. JOHNSON: Yes, I think that is very essential to this discussion. What we have tried to make clear, and will make clear in other presentations, is that the surrogates are based on as much information as we have had prior to the mid-'70's, when this regulatory mechanism was invoked, until now. There still is not a body of evidence, while we are asking you to assess that body of evidence and whether you think that compels us to change. So, there are standards in place and we think that those standards are based on the information that has been available to this point. At this point we are reconsidering the standards and we do think, and we are suggesting to the committee that any change in the standards should

be data driven.

DR. WOOD: Just to summarize, so what you are saying is that you don't want the committee particularly to consider the quality of the data supporting the standards; you want the committee to consider the quality of the data supporting a change in the standards.

DR. JOHNSON: Well, I think it is both but our concentration is really on the latter part of that.

DR. WOOD: All right, thanks. The next speaker will be Tia Frazier.

Regulation History of Healthcare Antiseptic Drug Products

MS. FRAZIER: Good morning.

[Slide]

I am Tia Frazier, and I am a project manager in the OTC Division, and I will briefly review the regulatory history of the monograph for OTC healthcare antiseptic drug products.

[Slide]

The monograph includes both consumer and

professional use products. Today we are addressing issues related to the professional use products included in the monograph, which we call the healthcare antiseptics. I will start first by defining the healthcare antiseptics. There are three recognized uses, that Susan has already told you about, included in the tentative final monograph. These are patient preoperative skin preparations used to cleanse patient skin prior to surgery; surgical scrubs which are used by operating room personnel prior to performing surgery; and healthcare personnel handwashes which are the soaps and leave-on products that are used by all personnel in healthcare settings prior to contact with patients.

[Slide]

We have two different mechanisms for regulating OTC healthcare antiseptics. Companies can submit new drug applications, which we call NDAs, for specific drug products to the FDA. Data provided in NDAs remains confidential. The second mechanism that we have for regulating these

products is the OTC drug monograph review process. Products submitted to the monograph review are judged on the safety and efficacy of their individual active ingredients. The data review for monograph drug products is public.

[Slide]

Just to add to this brief description, I will also tell you that the OTC drug monograph review began in 1972. At that time, and for some years later, the agency made determinations about the safety and efficacy of over 200,000 OTC products that were on the market at that time. We have reviewed 700 active ingredients in 26 therapeutic categories with the help of expert panels.

[Slide]

The advisory review panel reviewed and made recommendations on ingredients and products to further the development of a drug monograph. FDA then categorizes ingredients considered in the monograph review according to their safety and effectiveness for a particular use described in the

review. I won't say much more about how we categorize and evaluate ingredients since the focus of today's meeting is on the effectiveness criteria that we use to evaluate this particular group of professional use products. The OTC review panel's recommendations are then published in an advance notice of proposed rule-making, or ANPR.

[Slide]

After the ANPR is published we consider public comments as we develop a tentative final monograph, or TFM. A TFM is FDA's proposed monograph.

[Slide]

FDA usually receives more data and public comments on any TFM that we publish. Typically, we publish a final monograph after a tentative final monograph. In this case, we published a second tentative final monograph in 1994 after the first, which was published in 1978.

[Slide]

We, at FDA, have the current view that antiseptics do play a pivotal role in the practice

of infection control today. We operate from the presumption that antiseptics can decrease the number of organisms on the surface of the skin and this probably reduces the spread and development of nosocomial infections.

Based on this presumption, we adopted surrogate endpoints, measurements of log reductions on the skin surface that are intended to indirectly measure the effectiveness of antiseptics that we regulate. This is the reason that FDA and the European regulatory bodies selected this particular surrogate endpoint, the reduction of the organisms on the skin surface, to evaluate the effectiveness of these products.

[Slide]

The advisory review panel recommended in 1974 that we use surrogate endpoints to measure antiseptic effectiveness. To date, unfortunately, we still have not figured out how to design a clinical study that can measure the contribution of an antiseptic in reducing the likelihood of contracting or spreading nosocomial infection.

With any luck, today Dr. Pearson will explain later why designing studies like this is so difficult.

[Slide]

So, now I am going to go into the history of the monograph as it relates to the surrogate endpoints. The first defined surrogate endpoint for patient preoperative skin preparations appears in our 1974 ANPR. It was also incorporated in the first tentative final monograph which, I said, was published in 1978. Then the panel recommended a 3-log reduction in organisms on the surface of the skin as the requirement for patient preoperative skin preparation. At that time, NDA products were often approved for patient preoperative skin preparation indications based on their ability to meet a 3-log reduction and the monograph simply adopted this commonly used NDA standard.

It is important to realize that the effectiveness criteria used today to evaluate products marketed under the monograph are really based on the effectiveness criteria often applied to NDA products. NDAs, of course, can be approved

with alternate clinical endpoints and are not necessarily bound by the monograph standards.

[Slide]

Moving on to the surgical hand scrub criteria, the history on this is that Hibiclens is an NDA product that was approved in 1975 based on a new surrogate model developed to evaluate surgical scrubs. FDA incorporated the effectiveness criteria applied to Hibiclens surgical scrub into the developing antiseptic monograph. These criteria were published in our second tentative final monograph, on June 17, 1994.

Hibiclens is often included as a positive or active control in testing designs for antiseptic products. Because these are laboratory tests, companies are required to include a positive control arm using an approved product like Hibiclens to ensure that the tests are conducted correctly.

[Slide]

The current 3-log reduction criteria proposed for healthcare personnel handwashes in the

second tentative final monograph was based on FDA's evolving understanding of what the NDA products under review at that time could achieve.

[Slide]

As I have said before, this monograph is unusual because there are two tentative final monographs associated with it. In 1994 we elected to publish a second tentative final monograph rather than a final monograph to allow for public comment on the new testing requirements. The current proposed testing requires in vitro studies of the product spectrum and kinetics of antimicrobial activity and of the potential for the development of resistance. We also require in vivo studies of effectiveness under conditions that we think simulate how the product is actually used in that healthcare setting.

Another unusual aspect of this monograph is that it requires in vitro and in vivo testing not only for the approval of new products but also for the approval of new formulations. We require this testing to be done because changes in the

inactive ingredients or dosage forms can affect the product's effectiveness.

[Slide]

Products are required to meet key attributes important to their performance in healthcare settings. We state that a healthcare personnel handwash should be persistent if possible. We would like it to be non-irritating, fast acting and be able to kill a broad spectrum of organisms as well.

Persistence, or the ability to have a residual effect for some time after the product is used, is also an attribute that we would want a surgical scrub or a patient preoperative skin preparation to have as well.

[Slide]

We have had two prior public discussions about these effectiveness criteria. We discussed performance testing at an advisory committee meeting in 1998. This was a general discussion only and we did not present questions for the committee to vote on. Then in 1999 we held a

public feedback meeting to hear the industry coalition present an alternative model or framework for evaluating antiseptics. Dr. Jackson will cover the effectiveness criteria proposed by this industry coalition in her presentation that follows mine.

[Slide]

I think everyone here today would agree that it is critical that FDA ensures it uses the right criteria to evaluate antiseptic products. There are many dangers we can imagine might occur if we allow ineffective products to be sold and used in hospitals. We need these products to work. The OTC and anti-infective divisions admit that the effectiveness criteria we currently use are not based on data from clinical studies. We recognize this as a limitation of our current standards.

The divisions recently reviewed available scientific data on topical antiseptic products used in healthcare settings. We searched for data that could be used to support effectiveness standards for this class of products. Our review of more

than 1,000 studies submitted by industry and picked up through our own literature search is included in the committee background packages. Dr. Steven Osborne will present the results of his review and evaluation of a section of those references that address clinical benefit later on this morning.

[Slide]

The monograph for OTC healthcare antiseptic drug products is in the tentative final monograph or proposed rule stage. We are in the process of writing a final rule, and we need your recommendations on what the effectiveness criteria should be in order to finalize this monograph.

Now I would like to introduce my colleague, Dr. Michelle Jackson, who is a microbiology reviewer in the Division of Over-the-Counter Drug Products. She will review the testing methodologies used to evaluate these products.

Testing of Healthcare Antiseptic Drug Products

[Slide]

DR. JACKSON: My talk will focus on the

testing criteria for healthcare antimicrobial drug products, and currently the development and standardization of protocols regarding the testing criteria for healthcare antiseptic drug products are based on earlier NDA review process.

[Slide]

My presentation will discuss where we are with the proposed monograph requirements in regards to clinical simulation testing procedures for healthcare personnel handwash, surgical hand scrub and patient preoperative skin preparation, and the use of surrogate endpoints, also referred to as log reductions, with the three healthcare professional products. Then I will go over the industry coalition's position of wanting to use alternative criteria. [Slide]

During the early stages of the antiseptic NDA review process standardized protocols did not exist. However, the agency requires standardized and reproducible methods, therefore, as the NDA review process evolved clinical protocols used throughout the NDA review process also evolved into

protocols now recommended in the tentative final monograph.

So, what makes a good clinical simulation test method? It should simulate as close as possible the actual use conditions. Ideally, clinical simulations should include design characteristics such as test product, also referred to as final formulation; the test product contains the active antimicrobial agent; a vehicle control arm is the test product without the active antimicrobial agent and vehicle, and negative control that shows how much contribution of reduction is due to just the mechanical action of washing the hands.

A current trial design in TFM does not recommend inclusion of a vehicle for healthcare personnel handwash and patient preoperative testing. The active control arm is also referred to as the positive or internal control. The active control is used to assess the reproducibility of the clinical simulation studies and also used to validate the study. This standard is usually a

chlorhexidine gluconate containing product. Clinical simulations should also measure the desired product performance. This simulation testing generates the surrogate endpoints and it should also be reproducible.

I will briefly go over the three testing criteria for healthcare personnel handwash, surgical hand scrub and patient preoperative skin testing.

[Slide]

For healthcare personnel handwash, the label indicated use is handwash to help reduce bacteria that potentially can cause disease. The products are used by healthcare professionals on a daily basis up for to 50 handwashes per day. The testing process predicts the reduction of organisms that may be achieved by washing the hands after handling contaminated objects or caring for patients. Here we are focused on the removal of transient organisms. The testing process is designed for frequent use and it measures the reduction of transient organisms after a single use

or multiple uses to initial baseline level.

The studies are designed to demonstrate a cumulative effect of an antiseptic, meaning that the product gets better and better in reducing the bacterial load on the hands. Thus, the products are considered broad spectrum, fast acting and, if possible, persistent. The TFM surrogate endpoints propose a 2-log reduction for the first wash and a 3-log reduction for the 10th wash.

[Slide]

For the inclusion criteria subjects participating in the studies must be between the ages of 18-69, generally in good health, and have no clinical evidence of dermatosis, open wounds, hangnails or other skin disorders.

The subjects are excluded if they have been diagnosed with having medical conditions such as diabetes, hepatitis, or having an immune compromised system, subjects having any sensitivity to antimicrobial products, pregnant or nursing women also would be excluded from participating in a study.

For the healthcare personnel handwash there is a one-week washout period where subjects are instructed to use a non-antimicrobial product,

such as soaps, deodorant and shampoos, and avoid bathing in chlorinated pools and hot tubs.

[Slide]

The outline of the test procedure includes a test practice wash using bland soap. This basically removes any oils and dirt from the hands, and the bacteria counts are compared to the baseline counts. The hands are contaminated with Serratia marcescens and immediately sampled, and the baseline is determining the number of organisms on the surface of the skin prior to using an aseptic product.

The handwashing schedule involves ten washes performed on one day. At the first wash the hands are contaminated and washed with the test product. The hands are then sampled for microbial counts. Eight additional washes are performed, and at the tenth wash the hands are sampled for microbial counts and the product must achieve a

specific log reduction after the first and tenth washes. The repetitive hand washing aspect of the study design is intended to mimic the repeated use of a product in hospitals. The repetitive washing is also used to measure the cumulative effect, and cumulative effect is a progressive decrease in the number of microorganisms recovered following the repeated application of the test product.

[Slide]

Once the hand washing procedure is completed, the subject's hands are decontaminated by sanitizing the hands with 70 percent alcohol. The purpose of this is to destroy any residual Serratia marcescens left on the skin. Typical handwashing procedures involve contaminating the hands with a microorganism, Serratia marcescens. The hands are rubbed together for 45 seconds, and the hands are held away from the body and allowed to dry for a few minutes.

[Slide]

Once the hands are dry, a specific amount of test product is dispensed into the cupped hands

and the next step is to lather and wash all over the surface of the hands and above the wrists. After the completion of the wash, the hands and forearms are rinsed under regulated tap water with a temperature of 40 degrees Celsius for 30 seconds.

[Slide]

The hands are then placed in plastic bags and sampling fluid is added to the bag containing neutralizers. Neutralizers are reagents that stop the antimicrobial reaction. Sampling should occur within five minutes after each wash. The bags are tightly secured above the wrist with a strap. The hands are massaged for one minute, paying particular attention to the fingers and underneath the nails.

[Slide]

An aliquot of the sampling fluid is aseptically withdrawn from the bag and transferred immediately to dilution tubes. The microbial count determination is performed by surface plating and this is done within 30 minutes of sampling. The plates are incubated for two days at 30 degrees

Celsius.

[Slide]

This diagram depicts the colony forming units, CFUs, from two dilution plates. CFUs are then converted into log counts. Serratia marcescens produces a red pigment color for easy identification, and it distinguishes itself from the normal flora of the hands that appear white or yellowish on agar plates. Here, I want to emphasize that we are just counting bacteria.

[Slide]

Here the industry coalition suggest a 1.5 log reduction for the first wash, and suggest eliminating the tenth wash. We require the test product to show a cumulative effect, that is an evaluable attribute, that shows a progressive decrease in the number of organisms recovered following repeated application of a test product.

[Slide]

For surgical hand scrub the indication use is to significantly reduce the number of organisms on the skin prior to surgery. These products are

used to reduce the resident and eliminate the transient flora of the hands of surgeons and surgical personnel, thus reducing the incidence of post-surgical site infection.

The testing process is designed to measure the immediate and persistent reduction of resident organisms after a single or repetitive treatment. Here there is no artificial contamination of the hands, and the testing of the surgical hand scrub involves multiple test product use and repeated measurements of the bacterial reduction. These antiseptics are considered broad spectrum, fast acting and persistent. The TFM surrogate endpoints propose a 1-log on day 1 for the first wash; 2-log on day 2 at the second wash; and 3-log on day 5 at the 11th wash.

[Slide]

The subjects are selected through the inclusion/exclusion criteria for surgical hand scrub testing. A 14-day or 2-week washout period is required. Soon after the washout period the baseline counts are determined, and they are

sampled two times, first on day one and the second estimate includes one of the three options. On day 3 and 5, 5 and 7, or 3 and 7.

Subjects with a baseline greater than or equal to 5 logs after the first and second baseline estimates will qualify for the study testing period. So, no sooner than 12 hours and no longer than 4 days after completion of the baseline determination subjects perform the initial scrub with the test product. The surgical hand scrub testing requires a total of 11 scrub washes over a 5-day period. The sampling occurs on day 1, day 2 and day 5.

The reason we test 5 days is that the procedure mimics typical usage and permits the determination of both immediate and long-term bacterial reduction. Each day the antimicrobial soap is used it produces a greater effect due to the persistence of minute residues left from the previous scrub. This effect is called cumulative effect, and that is the reason why we test for 5 days.

[Slide]

An amount of the test product is dispensed according to the manufacturer's labeling

instructions. The soap is distributed all over the hands and two-thirds of the forearms.

[Slide]

The hands are then scrubbed according to the manufacturer's directions, and if no directions are provided the TFM requires two five-minute scrub procedures. A scrub brush is used to scrub the hands including the nails, the fingers, and interdigital spaces of the hands.

[Slide]

A lab technician will don sampling gloves on the subjects. One-third of the hands in a treatment group is sampled immediately. The gloves remain on the test subjects' hands for either three hours or six hours prior to sampling. Enumeration of bacterial flora three hours after the scrub is conducted in order to demonstrate continued effectiveness of the product during the time required for a surgical setting. The enumeration

of bacterial flora six hours after the scrub is conducted to demonstrate the suppression of bacterial counts over a period of time chosen as representing the maximum duration of most surgical procedures, that is, on average most surgeries will not last greater than six hours and, if so, surgeons usually rescrub.

[Slide]

A specified amount of sampling fluid then is added to the glove pan, and the gloves are fastened securely above the wrist and strapped, and the hands are then massaged for one minute, paying particular attention underneath the nails.

[Slide]

An aliquot of the sampling fluid is aseptically withdrawn from the glove and transferred immediately to dilution tubes containing neutralizers. A microbial count determination is performed by surface plating, and this is done within 30 minutes of sampling. The plates are incubated for two days at 30 degrees Celsius.

[Slide]

Here the industry coalition agrees with the 1-log reduction for the first wash. They

suggest eliminating the second and 11th wash. They suggest that persistence of antimicrobial activity should not be a requirement for surgical hand scrub. We require an assessment of persistent activity in case there is a tear in the surgeon's glove, and it is assumed that the persistent effect will prevent the multiplication of resident flora on the gloved hand, thus preventing contamination of the surgical field.

[Slide]

For the patient preoperative skin preparation or surgical prep labeled for the indicated use helps reduce bacteria that potentially can cause skin infection. These antiseptic products must be fast acting, broad spectrum and persistent and, statistically reduce the number of organisms on intact skin. They are designed for use by healthcare professionals to prep the patient's skin prior to invasive surgery

or prior to injection. These indications, however, do not cover more specific indications such as catheter insertions and open wounds.

The testing process measures the immediate and persistent reduction of resident bacteria after a single treatment. The TFM surrogate endpoint proposed a 1-log reduction for pre-injection; 2-log for the abdomen or dry site; and 3-log for the groin or moist site area.

[Slide]

The subjects are selected through the inclusion/exclusion criteria for patient preop testing. A 14-day washout period is required, and no bathing 24 hours prior to the baseline screening. We want to try to obtain a high bacterial count for the baseline. The TFM recommends the baseline screening counts for pre-injection to be greater than or equal to 3 logs. The TFM recommends that baseline screening counts for the common surgical sites for both dry and moist site areas, and the sites are to present bacterial populations large enough to allow the

demonstration of bacterial reduction for up to 2 logs centimeters squared for the abdomen sites and up to 3 logs centimeters squared on the groin sites.

[Slide]

For the abdominal site testing a 5 X 5 treatment site area is marked on the skin using a permanent marker. The template is divided into four quadrants for baseline, 10 minutes, 30 minutes and 6 hours sampling.

[Slide]

The baseline sampling is performed using the cylinder sampling technique. A sterile scrubbing cup is held firmly against the skin over the site to be sampled. The scrub solution containing neutralizers is placed into the cup and scrubbed with moderate pressure for one minute using a sterile rubber-tipped spatula. This procedure is also used for sampling for the treatment site.

[Slide]

The application of the prep formulation is

applied to the testing area. For 30-minute and 6-hour sampling sites a sterile gauze is placed over the prep area to help prevent microbial contamination. The gauze pad is held in place by the sterile teeth dressing.

[Slide]

The treatment samples are taken from the site areas using the cylinder sampling technique. A similar procedure is also used for testing the groin site area.

[Slide]

Here the industry coalition agrees with the 1-log reduction at the pre-injection site, and they suggested that only a 1-log reduction should be required for the abdomen site and a 6-hour persistent is not needed. For the groin site a 2-log reduction should be required and a 6-hour persistent is not needed.

[Slide]

FDA has received objections to the TFM proposed effectiveness criteria through comments in a citizen's petition. Industry contended that the

current performance criteria for healthcare antiseptics are overly stringent. They claim that two category ingredients, alcohol and iodine, and one NDA approved ingredient, CHD, cannot pass the current testing requirements. They claim that all antiseptic products only need to be effective after a single use, and they also do not want to meet the persistence requirement.

[Slide]

This table summarizes the bacterial log reduction in industry's proposal for the healthcare antiseptic compared to FDA current standards for final formulation for healthcare personal handwash, surgical hand scrub and patient preoperative skin preparation I just reviewed. Over the years the industry coalition has made several proposals for the revised effectiveness criteria.

For the healthcare personal handwash, it should be effective following a single use. A cumulative effect should not be a requirement. For surgical hand scrub, it should be effective following a single use and also a cumulative effect

should not be a requirement. And for patient preop, the pre-injection and abdomen dry site a 1-log reduction is suggested, and for a worst-case scenario such as the groin site area, it should need a 2-log reduction.

[Slide]

We are aware the surrogate endpoints lack the clinical validation of a test method and performance criteria. They do not measure the level of residual bacteria on the skin and virulence of the residual bacterial is not factored into the log reduction determination. We realize that we are just measuring the mean log reduction.

The criteria is based largely on earlier NDA performance and we have approved over 20 NDAs based on using surrogate endpoints. These criteria are consistently applied to monograph products and many NDAs. Industry has deviated from following the TFM in regards to variability in testing procedures such as scrub techniques and lab analysis, and it is not compared to vehicle or active control. We will later hear from Dr.

Valappil regarding improving statistical analysis that could be applied to the existing criteria.

[Slide]

Overall, it is impossible to compare the data across studies due to the vast differences and methodologies that were used, and other limitations such as the following: The majority of the studies were designed as product comparisons; studies were not designed to assess the product's ability to meet the TFM effectiveness criteria. There were significant variations in how the studies were conducted; different testing procedures were used; and neutralizer validation data were not generally provided. More than half the data submitted did not include neutralizers in the testing procedures, which can result in artificially high log reductions. Generally, sample sizes were small in the studies and there was a limited number of subjects included in the testing procedure. And, alcohol alone did not meet the 10th wash 3-log reduction. However, most were able to meet the 3-log reduction of the first wash. We are

currently evaluating the alcohol leave-ons and alcohol gel products.

[Slide]

This slide was included to show that other countries also use surrogate endpoints. The European performance criteria for handwash require that the test product mean log reduction factor should be greater than soap that has an average reduction log of 2.8. The performance criteria for hand rub require that the test product mean log reduction factor should be equal to or greater than 60 percent isopropyl alcohol that has an average reduction log of 4.6.

[Slide]

In summary, we measure bacterial log reduction and testing methodology for healthcare personnel handwash, surgical hand scrub and patient preop. These log reductions are used as surrogate endpoints to evaluate effectiveness. How should we analyze this data?

Later this morning we will hear from Dr. Valappil a presentation on statistical analysis for

healthcare and aseptic drug products. You will also hear from Dr. Steve Osborne who will discuss the relationship of these outcomes and corresponding reduction in the incidence of nosocomial infections in healthcare settings where the product use remains undefined.

[Slide]

We are aware of the limitations of these test methods, and we assume that the incidence of infections as related to current use of existing products and lowering these standards may increase the infection rates. We need research to validate these surrogates, and we need to have products on the market now and in the use of actionable criteria in the meantime. That concludes my presentation.

DR. WOOD: Mike, you approached me earlier about some confusion about the data. Do you want to comment on that at this stage?

DR. ALFANO: Yes, I have been advised that industry is not recommending removal of the 6-hour persistence requirement but, rather, the cumulative

effect requirements. Apparently, that came about because of some confusion over a table that the industry submitted.

DR. WOOD: Can you put slide 12 back up? Is that the one that we are talking about here, on page 6? Is that where the confusion is?

DR. ALFANO: Actually, it was brought to my attention versus the questions that we are to answer today, which is on the last page of the agenda.

DR. WOOD: I was just trying to clarify these slides. So, there is no confusion about what industry's position is on the slides? Is that right?

DR. ALFANO: That is correct.

DR. WOOD: Well, I think there is actually. Somebody seems to want to comment.

DR. FISCHLER: George Fischler, manager of microbiology for the Dowell Corporation, representing the STA-CTFA coalition. Yes, there is some confusion. On this slide, yes, where it says surgical hand scrub, there is an asterisk and

patient preoperative skin preparation, an asterisk. Industry has not recommended the removal of the 6-hour persistence criteria. The only criteria that we recommended approval for is the cumulative effect.

DR. WOOD: Okay. Well, let's come back to discussing that later. I am even more confused now but let's go on to the next speaker.

DR. JACKSON: The next speaker is John Powers. He is the lead medical officer in the Antimicrobial Drug Development and Resistance Division, and he will discuss the biological surrogate endpoints in the clinical trials of infectious disease.

Microbiological Surrogate Endpoints in Clinical Trials of Infectious Diseases

DR. POWERS: Thanks, Michelle.

[Slide]

Today I am going to discuss issues related to microbiological surrogate endpoints in clinical trials of infectious diseases. Some of the members of the Anti-Infective Drugs Advisory Committee

won't be surprised by any of this since this is an issue that has come up in infectious disease trials over and over again. So, I am going to try to discuss just some of the general points that have to do with selecting surrogate endpoints in these types of trials.

[Slide]

The first thing I am going to talk about is differentiating what we do in clinical practice and how one develops clinical practice guidelines with what one actually does in a clinical trial, and how those are very different situations. Then what I would like to do is define our terms and talk about what is an endpoint; define what a clinical endpoint and surrogate endpoints are and differentiate those from biomarkers. One of the things you will hear often, and probably we will make the mistake today, is using the term surrogate markers rather than surrogate endpoints, which is rather non-specific and causes some confusion.

Then we will talk about the utility of surrogates in clinical trials and differentiating

surrogate endpoints from surrogates as risk factors, which is an entirely different consideration. I will talk about some of the strengths and limitations of surrogate endpoints and then, finally, relate all of that information to the use of surrogates in the setting of topical antiseptics.

[Slide]

What we do in clinical practice is we are using drug products that are already proven to be safe and effective and, hopefully, we are not experimenting on our patients; we are using the products in a way where they are already shown to work.

In clinical practice we impose several interventions on patients and hope they get better. We are not really concerned with why they get better when we do all that stuff to them, only the fact that they get out of the bed and they leave the hospital cured. We develop treatment guidelines to help us describe the use of the products based on whatever the best available

evidence is, and a lot of current treatment guidelines actually put grades on the evidence where you will see A-1 all the way down to D that talk about whether it is from randomized, controlled trials versus observational evidence as well, but optimally these treatment guidelines are based on randomized, controlled trials. When that data is not available we oftentimes have to put things into these guidelines based on the best available evidence that we have.

The unfortunate thing is that sometimes these guidelines then become the reason for not getting the data from randomized, controlled trials because people will come to us and say the guidelines say this, therefore, you can't do a trial to evaluate it. And, that is probably not what the people who alter these guidelines actually are intending.

This differs from clinical trials which are experiments in human beings to determine if drug products are safe and effective. Clinical trials differ from clinical practice in that we are

using the scientific method. We are trying to hold as much as possible constant, except for the interventions, so that we can apply the outcomes to causality related to the interventions themselves, which is very, very different from clinical practice. So, how we do this is often to use concurrent controls which is something that we do not do in clinical practice. In clinical practice we look at what the patient is at baseline and compare what happens at the end. That is not what we do in clinical trials where we are comparing what happens at the end in patients who receive the test product versus a control.

These clinical trials are, hopefully, to provide the evidence for formulation of practice guidelines and, as I said, hopefully, it is not vice versa where the guidelines determine that we can or cannot do a clinical trial. But the big issue in clinical trials is that we need to determine some yardstick to determine if products are safe and effective. How are we going to measure those products to make that kind of

assessment? That is really what we are asking today.

And, the reason for this slide is to sort of outline the real question today. We are not questioning whether handwashing is important or whether handwashing should be done in clinical practice. What we are asking today is how do we develop a yardstick to determine which products are safe and effective to use in handwashing.

[Slide]

So, let's define some of the terms that we are going to use today. An endpoint is a measure of the effect of an intervention on an outcome, outcome being defined, for instance, as success or failure in a clinical trial in the treatment or prevention of a disease. Again, it is important to realize that what we are talking about here is a disease. We are not preventing someone getting an organism on their skin. What we are really trying to look at is does that prevention of getting an organism on the skin, in turn, result in prevention of disease.

But whenever we are picking an endpoint we have several questions that we have to address. The first one is what are we going to measure?

Obviously, this should be clinically relevant to the disease in question. We are not going to ask if your left earlobe hurts when we are trying to evaluate something that has to do with foot pain.

The next question is how to measure it? And, we should be able to measure differences between therapies, should they exist, and that gets to this issue of the yardstick and that we need to be able to differentiate effective from ineffective products.

The next issue is when do we actually measure it? If we apply a product and come back in two years and then try to determine if there are differences between the patients we are probably not going to see a whole lot in a non-lethal illness.

The next question is how much to measure, what magnitude of difference actually makes a difference to patients? A lot of this has to do

with sample size. We could take a product that is 99 percent effective and show that it is statistically different than a product that is 90 percent effective if we studied thousands and thousands of patients. So, it gets to the issue of clinical significance versus statistical significance.

Then, one of the big issues I am going to ask you to talk about today is when we get some results, how do we analyze those results so that we can logically draw conclusions from them?

[Slide]

This is a cartoon from the New Yorker, which sort of outlines the issue in choosing endpoints that are relevant to patients. Here there is a doctor who has just done an endoscopy on a miserable patient, and the doctor says congratulations, the endoscopy was negative; everything is perfectly all right. So, according to the surrogate endpoint of what the doctor saw on the endoscopy, the patient feels great but the patient is saying my symptoms bother me. I am

worried and concerned. I can't exercise; I can't eat. My whole life is affected. So, that gets to the difference between measuring a surrogate and measuring what the patient actually feels.

[Slide]

This seems sort of redundant but it is probably important to define what a disease actually is. In these terms we are talking about a constellation of signs and symptoms experienced by the patient. Although infectious diseases are caused by pathogenic organisms, those result in a host response and it is actually the host response that causes a lot of the symptoms that we see.

When we are talking about surrogates we often hear about Koch's postulates. Well, these fulfill Koch's postulates so the surrogate must work in the setting of an endpoint of a clinical trial. But Koch's postulates relate to proving the cause of a disease, that a pathogen actually causes that particular illness, and Koch's postulates were never designed to measure the effect of an intervention. It is very important in our

discussion today to separate out cause from effect which are two different considerations.

One of the issues we always talk about is that patients seek the care of clinicians because they have symptoms when they have a disease, not because of the presence of an organism. So, a patient may come and say, doctor, I have this terrible cough I can't get rid of it. They don't come in and say, doctor, I have mycoplasma in my respiratory tract. Although that may be the cause of it, the reason patients come to see us is for relief of symptoms.

In prevention trials, on the other hand, we are actually seeking to prevent those symptoms from ever occurring, but still here we are talking about the relevant endpoints being those actual symptoms that patients may encounter.

[Slide]

So, what is the difference between clinical endpoints and surrogate endpoints? We are so used to using surrogates that sometimes we call things clinical endpoints that are, in fact,

surrogates. The definition of a clinical endpoint is actually fairly simple. It is measures of how the patient feels, functions or survives, and a simple way to think of it is anything that measures something other than that is a surrogate endpoint. For instance, clinical endpoints would be measures of mortality or resolution or prevention of symptoms of a disease.

On the other hand, surrogate endpoints are laboratory measurements or physical signs used as a substitute for a clinical endpoint. Fever is a surrogate endpoint. Fever does not necessarily measure how the patient feels. Although fever may make the person feel terrible, what we really want to measure is the person feeling terrible not what the level of the temperature is but we are so used to using this in infectious disease trials. But other things like culture results, which we are going to talk a lot about today, chest x-rays, histology or even data like pharmacokinetic information are all surrogate endpoints and need to be correlated with what is actually clinically

happening to the patient.

The important part here, as discussed at NIH Biomarkers Definition Working Group, published in 2001, is that surrogate endpoints by themselves do not confer direct clinical benefit to the patient and we need to make that link. This is also reiterated in the International Conference on Harmonization, ICH E9 document. The International Conference on Harmonization is a group consisting of U.S., Japanese, European regulators and members of the pharmaceutical industry.

[Slide]

So, how do we differentiate biomarkers from surrogate endpoints? Biomarkers are any set of analytical tools that are used to assess biological parameters so it is a big, broad category. Biomarkers are useful for many other purposes other than surrogate endpoints in trials. This is why the term surrogate marker isn't really very helpful to us because we can use these biomarkers for any number of things. One may be as a diagnostic tool. We can use the test as

inclusion criteria to define the disease based on the presence of organisms. Differentiating diagnosis from endpoint is a very, very important process. As members of our Anti-Infective Drugs Advisory Committee that are here will tell you, we have had several advisory committees for instance addressing acute otitis media in children and acute bacterial sinusitis in children and adults where we have tried to make the distinction between needing microbiologic data to diagnose that the person actually has the disease, but how useful it is as an endpoint is an entirely different consideration.

We can also use biomarkers to describe the mechanism of action of the drug and the effect on the organisms of an antibacterial or antiviral product is really the mechanism by which it achieves its effect, not necessarily the goal of therapy alone. We have certainly been told by a number of sponsors--the direct quote, all antibiotics do is affect organisms. Well, that is true but that is the mechanism by which they do what they do, not the goal of why we give them to

patients in the first place.

The third thing is that biomarkers can be a risk factor for acquiring the disease. For instance, we know that colonization with a particular organism is a risk factor for getting an infection. That doesn't mean that risk factors end up being the same thing as an endpoint. Also, some of these things can be risk factors for outcome. They can indicate disease prognosis and how poorly or well the patient is going to do. For instance, HIV viral load and CD4 counts in HIV--we can look at those to actually predict how a patient is going to do down the line. Then, finally, biomarkers can be used as surrogate endpoints, which are different from the previous four things we talked about.

[Slide]

The word surrogate comes from the Latin root surrogatus, which means to choose in place of another, or to substitute or put in place of another. So, what we are doing with a surrogate endpoint is actually substituting microbiologic outcomes in patients for clinical outcomes. One of

the problems in looking at this is that investigators have looked at people only who have failed and then tried to relate clinical and microbiological outcomes in only the failures. But we need to look at these correlations both in people who succeed and people who fail, which is pivotal in these clinical trials to prove drug efficacy.

[Slide]

Surrogate endpoints are very useful. They can be used in early drug development as proof of principle that the drug has some biological activity, and they can be used in selecting candidates to go on and study in future phase 3 trials. They are also useful in phase 3 trials when the surrogate endpoint can be measured sooner in time than the clinical endpoint. The obvious example of this is HIV trials, which I will go into in a little more detail.

When the clinical endpoint events are more rare it allows us to complete a trial with a smaller sample size. In other words, if the effect

on the surrogate endpoint is quite large and the effect on the clinical endpoint is small, we can do a trial with a smaller amount of patients in a shorter amount of time. Of course, this is all predicated on knowing that the surrogate actually predicts clinical outcomes.

Some examples of where the agency has allowed surrogates and they have been used successfully are things like lowering cholesterol which, in turn, has been shown to prevent cardiovascular disease; lowering blood pressure to prevent cardiovascular disease; and perhaps the best example is suppression of HIV viral load as a surrogate endpoint in the prevention of either AIDS-defining events or death in the treatment of HIV and AIDS.

[Slide]

In this example what we see is a three-dimensional graph. On the right-hand side there are CD4 counts which actually are predictors of the host's immune response. On the other axis is the viral load, or HIV RNA concentration. On

the upward axis there is the three-year probability of patients progressing to AIDS. You can see from this that as the person's CD4 count declines and as the HIV viral load goes up, the risk of developing AIDS-defining events and death also goes up. So, both HIV viral load and CD4 counts are predictors of what is going to happen to the patient independently.

The interesting thing about this is that this is measuring the organism but CD4 count is also measuring the host's immune response. HIV is very unique in that the virus itself blunts the host's immune response so one of the things that complicates the measurement of surrogates is that measuring the surrogate itself often doesn't measure what is happening to the person. So, viral load is very unique in that the virus itself knocks out the immune response and takes that piece out of the equation.

[Slide]

So, HIV viral load and CD4 counts are also a good example of the difference between risk

factors and endpoints. Both HIV viral load and CD4 counts are risk factors for disease progression to HIV and AIDS, as I showed you on the previous slide, however, only HIV viral load functions well as a surrogate endpoint, much better than CD4 count does in clinical trials.

Seven of eight trials with a positive effect on CD4 count also showed a positive effect on progression to AIDS or death. But the effect in 6/8 trials that had a positive effect on CD4 count also showed a negative effect on AIDS progression or death. This again gets back to the issue that you cannot cherry-pick which studies you like. You need to look at both success and failure of the surrogate to be able to get an overall assessment of what is going on here. If we only looked at these studies we would think that CD4 count was great as a surrogate endpoint.

This also gets to the issue that how you use the surrogate is very important. It may be that CD4 count would function as a decent surrogate endpoint if we followed patients for longer periods

of time than we follow the viral load because it just may be that the CD4 count may not change fast enough over the time that we measure it in a clinical trial to be very useful. But if we measured it for longer, that may be a different story.

[Slide]

What are some of the strengths and limitations then of evaluating surrogates? Part of this is the logic string we go through as related here to topical antiseptic products. We know colonization with organisms precedes infection and, therefore, the surrogate may be useful as a risk factor for disease. We know that these organisms can cause infection and result in a host response. So, the logic is that since the organisms cause infection, eliminating or decreasing the organisms should result in positive clinical outcomes for patients. This seems very logical. It seems very objective and reproducible. But the question is, is it correct?

This article by DiGruttola, and Dr.

Fleming is a co-author on this, talks about are we being misled in terms of looking at these surrogates? What we just did up here was an example of the old Arthur Conan Doyle Sherlock Holmes deductive reasoning. We worked backwards from the end and said, well, it must be caused by this. However, what we do in clinical trials is inductive reasoning. We start off with a hypothesis and we test the hypothesis. So, we need to test this logic to see if it is actually true. One of the seminal articles on surrogates was written by Prentice where he actually says that in a given clinical trial we need to test does the intervention have an effect on the clinical outcome and, in the same trial, does that intervention also have an effect on the surrogate so that we can link the two together?

[Slide]

Well, why may it be that an intervention having an effect on a surrogate which, in turn, has an effect on the clinical does not predict what actually happens to the patient? And there are

five potential reasons why this may happen.

The first is that there may be unmeasured harms caused by the intervention which actually are not picked up by just measuring the surrogate.

The second is that there may be unmeasured benefits, that the intervention actually does something good that is not measured by the surrogate and actually has a better clinical outcome than predicted by the surrogate.

The next issue is that there may be other pathways of disease that result in a clinical endpoint that have nothing to do with the intervention that you applied.

Finally, there are issues with how we measure the surrogate and how we measure the clinical endpoint. Let's go through each one of those one at a time.

[Slide]

As I said, surrogates may not take into account unmeasured harm and benefits. This gets to the issue of we cannot just look at whether a surrogate correlates with a clinical endpoint

because, even if there are these unmeasured harms and unmeasured benefits, there will still be an association between the surrogate endpoint and the clinical endpoint. It will be, however, that that association is not predicting the net clinical outcome in patients because it is not taking into account these other unmeasured benefits and harms.

It is not too hard to understand why this occurs because the body actually has a finite number of processes to accomplish the things it wants to accomplish. So, giving a drug product is still giving a foreign antigen to the body which may affect processes other than the ones that we actually intended to affect in the first place. We know that, for instance, in antimicrobial products what we are really trying to affect is the organism which, in turn, has a positive effect on the host. The reason why we get adverse events is that all of these products have some effect on the host that is unintended in terms of adverse events.

[Slide]

What are some examples of unmeasured

benefits? Well, there may be effects of the drug other than eradication of the organism. Actually, this is a misnomer. We constantly use this term "eradication" but what we really mean is that we have suppressed the organism to below a level of detection. If we think that we are actually sterilizing somebody's body, we really are fooling ourselves. There may be sub-inhibitory effects of antimicrobials on the organisms. Even though those organisms are present, they can't do what they normally do in terms of invading. It may be that we don't need to kill the organisms to actually have some effect on the ultimate outcome and, again, that may be because we are having other effects, other than killing, that do something to the organism. Then, again, there may be direct effects of the antimicrobials on the host immune system. These articles that I have shown up here are actually things that talk about the effect of antimicrobial products on white cell phagocytosis and other processes on the human immune system.

There also may be unmeasured harms in

terms of deleterious effects on the host that may promote infection. For instance in topical products, if a product actually would cause micro-breaks in the skin that would not be visible to either the infection or the patient that may allow more invasion of organisms to cause wound infections. We also may have replacement of one organism with another. We get rid of the one organism we are worried about and, nature abhors a vacuum, and something else comes in its place that is actually worse than what we got rid of. There may be other sources of infection, other than those affected by the drug.

[Slide]

Are there some examples of where we have seen this happen in the past? The answer is yes. This is why we have such pause when evaluating surrogates. For instance, last year the FDA approved rifaximin as a treatment for travelers diarrhea. If one evaluates the rate of negative cultures from the stool in rifaximin compared to placebo, there was no statistical difference

between the number of organisms at the end of treatment in the stool in patients who received the drug versus those who did not.

Regardless of that, there was still decreased time to resolution of diarrhea with rifaximin compared to placebo. You could say, well, that means rifaximin isn't acting as an antibacterial agent; it is doing something else, it is decreasing GI motility. Well, if that is the case, then why did rifaximin have an effect on some organisms like E. coli, but not on diarrhea caused by other organisms like Campylobacter? If it was just acting as a motility agent it should have equal effects on everything. So, perhaps this drug is doing something to the organisms other than killing them.

Other examples of unmeasured harms--well, a classical example of this is the dose escalation trial of clarithromycin that was studied at 500,000 and 2,000 mg for disease due to Mycobacterium avium-intracellulare in patients with AIDS. When we looked at that dose response, the higher doses

had higher rates of negative blood cultures for MAI. However, those higher doses also had higher mortality in terms of the clinical outcomes. So, a better microbiologic outcome actually resulted in a worse clinical outcome in this trial.

[Slide]

Are there also other pathways of disease that may be unaffected by the intervention? Do we have an example of that?

[Slide]

Well, several trials showed decreased rates of colonization in the nose with Staph. aureus with intranasal mupirocin. However, three trials now done in the last several years show that prevention of infections with mupirocin, the clinical outcome, was not lower in patients than placebo even though there was a dramatic effect in terms of negative cultures done from the nose with this particular product. One hypothesis for why this may not be effective is that Staph. aureus is on numerous sites on the body other than just your nose and we may not be affecting that just by

putting a product on one site in the body.

[Slide]

The next issue is with accuracy of how the surrogate is measured. One of the things that we constantly hear about surrogates is that they are reproducible. Well, reproducibility talks about precision, but the example you can think about here is how to differentiate precision from accuracy. If I take a bow and arrow and I shoot it at a target I can hit the same spot on the target all the time, but it may be way far away from where the bulls eye actually is. So, even though we are getting reproducibility, are we getting accuracy? Are we getting the correct inference? This has to do with what, when, how and the magnitude of what is measured for that particular surrogate.

[Slide]

The culture techniques that we use for bacteria are based on methodology actually from the late 1800's. We know that there is inherent error. For instance, if we take the exact same colony of organisms and measure it two separate times we can

get minimum inhibitory concentrations for a particular drug that are actually off by one or two tube dilutions jut by testing it a second time. So, we know that there is some inherent error here.

There are a lot of issues with microbiological outcomes. For instance, what is the patient population that we sample? What is the sampling technique that was used? What was the methodology used to get the culture? Actually, I see Al Sheldon sitting in the back. When he used to work for us he gave a great talk last year on diabetic foot infections where we talked about how superficial cultures from the foot may not tell us anything related to deeper cultures from the foot in diabetic infections, and that methodology is very important.

When is the culture performed? On therapy cultures may be very misleading because when we take a sample we are actually taking the antibiotic with it and putting it onto the culture plate as well, which may give false-negative cultures.

How often do we sample, and what is a win?

What is the criteria for classifying that this organism is there or not? Do we have an all or nothing approach that says bug present/bug not present? Or, do we so something like HIV viral load where we have a quantitative assessment of how much organism is present?

[Slide]

The quantitative assessment may be very important, as I show on this graph. On the bottom axis we have time where we can make a baseline measurement and on therapy measurement and what happens when a drug is gone after the study is over, compared to microbial load. If one patient starts out at a higher level than the other patient, they both may decrease simultaneously at exactly the same rate, but if we make an on therapy assessment this patient may still have a positive culture and this one does not just because we have gone below some level of detection of how many organisms we can actually detect. Does that mean that these two patients are really different? We don't know. It may just be a factor of how many

organisms we were actually able to detect. If we only looked at an on therapy assessment, that may not tell us what happens after the drug is removed from the body. In one patient the bugs may come roaring back because all we did was suppress them. In the other patient it may continue to decline and we get rid of the organism altogether.

[Slide]

One of the issues that I am sure we will talk about today is this issue of practicality, and practicality is in the eye of the beholder when it comes to clinical trials. People have said because it is difficult to measure the clinical endpoint we should just rely on surrogates, which is very difficult logic in terms of perhaps needing to do a better job of actually measuring clinical endpoints. An inaccurate measurement of clinical endpoints does not justify the use of unvalidated surrogates.

[Slide]

For example, there is a recent article, and there has been an ongoing debate in the

Clinical Infectious Disease journal about the utility of catheter tip decolonization which, in this study, are claimed to be validated as a surrogate endpoint for clinical trials in prevention of catheter-related bloodstream infections based on the correlation of the two endpoints. What they did, however, in these trials is they defined a bloodstream infection in some of these trials as a positive blood culture and a positive culture of a catheter tip. So, this correlation is highly dependent upon the definition of the clinical endpoint.

Dr. David Patterson, from the University of Pittsburgh, wrote in about one of these studies and said, residual antimicrobial activity in the removed catheter sufficient to prevent growth from the cultured catheter segments would substantially reduce the apparent rate of catheter-related bloodstream infections--and I put the emphasis on there--could it be that use of these coated catheters impregnated with antibiotics prevents growth from catheters in the microbiology

laboratory but does not eliminate the clinical syndrome of catheter-related bloodstream infection?

So, a more rational use of an endpoint here would be all people that have positive blood cultures and symptoms of a clinical infection, not just those who have to have a positive catheter tip because that is circular reasoning.

One of the issues we always get into at the FDA is what gets published is all the successes, and people will look at those and say, look, there is this great correlation. What is missing, and there has also been a lot in The New York Times recently, is about negative trials. What is missing is the data the FDA sits on showing where those surrogates did not work. We have had several examples now, both in catheter tip decolonization and in products that are actually put on topically around the catheter site, where they had a dramatic effect on decolonizing the catheter and no effect at all relative to placebo in preventing bloodstream infections. I cannot enlighten you anymore than that because this is

proprietary information and we can't share it, but the interesting thing sitting at the FDA is you always wish that you could talk about the negative examples but, unfortunately, we can't share those.

[Slide]

One of the other issues with correlating a surrogate is how well does it actually predict outcomes? A perfect correlation would be a slope of 1 in terms of evaluating the surrogate related to clinical success so an 80 percent success rate with a surrogate would result in an 80 percent success rate in the clinical outcomes. But we don't expect that to happen, especially in prevention trials where we know that a good number of people on these trials will achieve no benefit from the product. So, what we want to look at is what is the actual correlation between the surrogate and the clinical outcome.

[Slide]

The other thing that is very important is that the correlation may differ from drug class to drug class or from drug product to drug product,

and this may actually be highly misleading in terms of what we actually measure. For instance, let's take drug A and drug B that have two different correlations in terms of the clinical and the surrogate. If we did then a measure of drug A and drug B in terms of the surrogate, it appears here that drug B is better than drug A in terms of the outcome with the surrogate. But if these two slopes of the correlation are different what actually is misleading is that in reality drug A is actually better than drug B in terms of clinical success so the surrogate actually flip-flops these and misleads us in terms of telling us why would these slopes be different.

That gets back to the five things we actually talked about. Unmeasured harms, unmeasured benefits and those other things may be why these products have different correlations. We actually did this with otitis media and showed that the spread of lines here actually goes from 0.4 all the way down to 0.1 for various different drug products. So, saying that this won't occur--we

have actually seen places where this correlation is actually all over the map for various drug products.

[Slide]

Finally, there are regulatory issues with surrogate endpoints. Traditional approval is based on surrogate endpoints only in cases where the endpoint is already validated to predict clinical benefit. However, there is an accelerated approval clause in the Code of Federal Regulations based on surrogate endpoints for serious and life-threatening diseases, otherwise known as Subpart H. This is where a surrogate endpoint is reasonably likely to predict clinical outcome. However, this part of the Code of Federal Regulations requires confirmatory post-approval trials based on the clinical endpoint to prove that what we saw with the surrogate is actually true.

The important thing to note today is that this clause actually came out in the mid-1990's and what we are talking about today is a monograph that started out in the early 1970's. So, if you ask

the question, well, why doesn't the monograph jive with what we are saying up here, it is because we are talking about something that happened 20-30 years before this regulation.

[Slide]

Let's relate all of the stuff we just talked about with surrogates to the issues related to topical antiseptics. Are there some potentials for unmeasured harms with topical antiseptics? Well, we may have unintended effects on microscopic breakage in the skin which may actually result in a greater clinical infection rate. We know this can happen, for instance, in trials that examine peri-operative shaving. This trial by Seropian, done in the American Journal of Surgery in 1971, actually showed a 5.6 percent rate of postop infection with shaving compared to a 0.6 percent rate without shaving. So, we know that there can be unintended effects.

If you go back and look at the hypothesis of that trial, it was exactly what we are trying to say today, clipping hair off may decrease the

amount of bacteria near the wound and, therefore, should result in a decrease in infections. It didn't; it did the exact opposite because of unintended harms that they didn't think about until after the trial was done. It is always fascinating to see how someone's hypothesis changes after the actual results come out.

Also, the effects on common pathogens may be less than that on the marker organisms on the skin. Michelle Jackson showed you that what we are measuring here is resident microbial flora in two of the three indications and we are contaminating people with Serratia marcescens in another. Serratia marcescens is not a common cause of skin infection so the question is does predicting an effect on Serratia tell us anything about staph., strep., E. coli, enterococci and the other common causes of infection?

Also, there is this issue of are we selecting resistance to systemic antimicrobials by using these topical antibiotic products? This really is something that deserves its own whole

discussion, but there is some evidence at least in the test tube that there may be afflux pumps which confer resistance to both topical products and to the systemic antimicrobials simultaneously, at least in E. coli and Pseudomonas. People have questioned what is the clinical relevance of that but that really is the question, isn't it? Once again, it is how does that surrogate predict what is going to happen clinically? I always think it is fascinating when you don't want to use a surrogate, all of a sudden it is not relevant. When you do want to use a surrogate, we will accept everything we want to believe about it.

So, can there be unintended benefits? Well, it may be that some of these products have positive effects other than those on the organisms. It does something to the host immune system that actually results in a decreased infection rate, more than we would predict by what it does to the bug. Also, could the effects on common pathogens, like staph. or strep. be greater than on something like Serratia? So, it may be a better benefit than

what we think.

[Slide]

Are there other mechanisms not affected by the intervention? Well, at least in terms of patient preop, for that indication we can look at a study that was done by Brown et al. in 1989 at the University of Virginia. The data that we are obtaining from this surrogate is really from the most superficial layers of the stratum corneum of the epidermis.

[Slide]

Here is an anatomical picture of the skin. What you see here is that the top 30 layers of the skin are this dead, keratinized layer called the stratum corneum of the epidermis. What is down here is the stratum germinativum where these cells come from. The cells die off. They become highly keratinized at the stratum granulosum layer which forms a barrier between this and the stratum corneum. What we are measuring in these trials is what is way up here.

[Slide]

So, what is way up there is right here on this graph. This is actually from the CDC guidelines on prevention of surgical infections.

What we are worried about is infections here, here, here and here. So, the real question is does doing something up here do something down here in terms of affecting the organisms?

[Slide]

This group in Virginia actually did a very elegant experiment with a methodology that was developed by Pincus in 1952. What they did was they took regular old cellophane tape and they showed that by putting cellophane tape and stripping it off the skin you can take one layer of that stratum corneum off at a time. They evaluated this in 12 different sites on the body, and they showed that these 12 different sites in the body had highly variable colony counts of organisms depending upon whether you are looking at the arm, the back or other sites.

They also showed that the number of colonies decreased over the top five layers of the

stratum corneum but then stabilized in the remaining 20 layers of the stratum corneum. So, there were more organisms up at the top than there were in the lower layers of the stratum corneum.

But then they did something very interesting. They took alcohol and decolonized the area that they had stripped, put a gauze pad over it and came back 18 hours later. They then did plasmid profiles on the coagulase-negative staphylococci that were there at the beginning of the experiment and there 18 hours later and saw identical plasmid profiles for those staphylococci.

So, they hypothesized that this indicates a reservoir for these organisms that may be below the stratum corneum, in the hair follicles and sebaceous glands of the dermis so where infection may come from is actually from the organisms that are lower down. This is one of the reasons why we give systemic antimicrobials as perioperative prophylaxis, trying to affect those organisms that may be down deeper in the dermis.

We also know that studies in perioperative

systemic antimicrobials show that if the antibiotic isn't around at this layer at the time you get operated on they will not be effective. For instance, you cannot give the antibiotic two seconds before you make the surgical cut because they will not affect the subsequent infection rate.

[Slide]

Then there are all the issues with measurement of the surrogate, which we are going to talk about today. Are we actually measuring the surrogate in a population that we are going to use it in? No, we are not. We are measuring healthy volunteers, not healthcare workers or patients.

As we already discussed, the organisms measured are not necessarily those that cause infection. Is the timing of these measurements relative to the disease process we are actually trying to prevent? That gets at this issue of do we need to get persistent effect or not; how long do we have to look for that; and how long should we look for it? For instance, we know that some patients may undergo prolonged surgery. Surgeries

may last hours and hours so an immediate effect is not the only thing we want to look at.

Are the conditions of testing the same as those that would be encountered in real-life situations? And, what happens with variations in the methodology? One of the things that is interesting at the FDA is that you will see people submit things that say I am using the such-and-such method approved by the CDC or the NIH. But it is a modified method. I always joke I am a modified millionaire movie star; I am just not a movie star and I don't have a million dollars. So, modifying the method--it is no longer the method. So, we need to take into account that changing the method, even if we have a valid surrogate, may actually change the correlations between the surrogate and the clinical outcomes.

The next question is what log reduction is clinically significant? And, how do we analyze those numbers obtained on log reductions? Dr. Thamban Valappil is going to go through a great talk that actually walks through some of these

issues with how do we analyze the numbers.

[Slide]

What is the data showing correlation of reduction of bacteria with a decrease in infection rates? Steve Osborne is going to go through our, believe me, exhaustive, over 1,000-paper literature search. You should have helped us out with this; that was a thrill!

What does the dose-response curve look like for infection rates and numbers of bacteria? Is it a threshold effect, or is it a continuous variable, and is it the same for all types of products?

[Slide]

What do I mean by dose response? Down on the bottom it should read numbers of bacteria on the skin, not change in numbers of bacteria. On the Y axis we have rates of infection. What we want to know is does the dose-response curve look like this? Sorry, this doesn't show up very well but it is a straight line. Or, does the dose-response curve look like this? The first

straight line is a continuous variable. The more organisms there are, the more infections patients get. The curved line is really a threshold effect that we talk about. At some certain level of bacteria people are more likely to get infected and below that level they are less likely to get infected.

Why is this important for us? Well, if we look at a linear correlation between numbers of bacteria and rates of infection, what we will see is that the decrease of the numbers of bacteria by this much will actually result in a corresponding decrease in the number of infections by some amount.

[Slide]

On the other hand, if it is a sigmoidal threshold type effect, what we will see is that that same, exact change in the number of bacteria if it is on the flat part of the curve results in very little change in infection. So, this gets to what does a 3-log reduction actually mean? If this is 10 7 and this is 104 that is a 3-log reduction but

we are on the flat part of the curve so there is very little effect on what happens to the patient. If we go from 10
4 to 101 that is a 3-log reduction too but if we are on the steep part of the curve that may be telling us something very, very different. So, where you start may be as important as what the delta change is, and we don't have any information to tell us what this dose response actually looks like.

[Slide]

What I would like to leave you with then is sort of the thought process we have had to go through for the last several months in terms of trying to look at this. The first question you have to ask is what kind of endpoint are you going to pick to evaluate these products? Are we going to pick a clinical endpoint or a surrogate endpoint? Ideally, there would be the data right here that links the clinical and the surrogate endpoint together, and Steve Osborne is going to talk about our attempts to actually make that kind of a link.

The second question is what are we actually going to measure? Let me get back to this issue of practicality. As I said earlier,

practicality ends up being in the eye of the beholder. One of the things you will hear about is that it takes more patients to do these clinical trials than it does to the surrogate endpoint trials.

Well, size is actually an issue but size really relates more to the time that it takes to do a trial which, let's be honest, relates to cost to do the trial. One of the questions you have to ask when you are getting into this debate is how much does it cost to do it wrong? How much does it cost the patients if we don't get this information and we don't actually know whether these products are effective? That side of the equation needs to be factored in as well.

The other issue that comes up is ethics. Ethics are only if you are denying somebody a proven effective treatment. What we are trying to evaluate here is are these things proven effective

or not, so we need to keep that in mind when we are discussing the ethics issue. When we talk about clinical trials the endpoint is very simple, it is infection in patients. On the other hand, with the surrogate we are looking at numbers of bacteria.

Then we need to talk about how do we design these studies and how do we define success. Well, the definition of success, again, with the clinical endpoint is much simpler actually. It is just the percent of patients that don't get an infection. However, when we talk about selecting an endpoint for a surrogate we have several decisions to make that Thamban is going to go through. Do we look at mean log reductions, median log reductions, the percent of subjects who meet some log reduction? And, where do you get this information from? Well, actually optimally it would be from a clinical trial that evaluated both of these things simultaneously.

Finally, how do we analyze the results that we get? Again, it is much simpler in a clinical trial. We just compare it with a

concurrent control. This is one of the issues when people point to the studies, and Steve is going to go through this in some detail, they say we already know these things work. There is no concurrent control. What these things are is quasi experimental studies where they took what we were doing last year and they applied something new in the hospital and said, look, my infection rate went down.

What that ignores is natural changes in baseline infection rates that may occur. Even though the trials say, well, we didn't do any other interventions on these patients, you know in the real world and, hopefully our AIDAC members can enlighten us on this, when you have an outbreak of some particular organism you do not do one intervention. You cohort patients together; you start using gowns and gloves on those people; you do a lot of other interventions that really call into question what was the cause of why the infection rate went down. Was it just related to the product that you used?

So, here we would make this comparison and either design these as superiority or non-inferiority trials, otherwise called

equivalence trials, that show that the product is no worse than something that is already out there.

On the other hand, there are a lot more complex decisions with a surrogate endpoint. Do we say that these things meet some threshold that we set? If so, where does that threshold come from? Where does the data come from to say? And, do we still need some comparison with a control given the variability in the method? Michelle Jackson showed you on one of her slides that at least that article in The Journal of Hospital Infection, based on the European methodology which is slightly different from that that is in the TFM, shows at least a 2 to 2.5 log drop with soap and water all by itself. So, do we need to look at how these things compare to some vehicle or another product? And, again, we have the choice of superiority or non-inferiority.

[Slide]

To conclude then, surrogate endpoints must

not only correlate with clinical outcomes but they must also take into account unmeasured harms and benefits; the methodology and uncertainties in measuring the surrogate; and the appropriate measurement of the clinical endpoint.

The clinical endpoint for efficacy of topical antiseptic products would be prevention of infections but actually the clinical design of these trials would vary depending upon whether we are talking about patient preop surgical hand scrubs or healthcare personnel handwash.

One of the things that I am sure we will hear about is what Semmelweis did in 1847 was he showed that medical students who went and examined corpses with their bare hands and then went and delivered babies--there was actually a higher rate of death in the mothers who had their babies delivered by these medical students than the midwives who were spared the odious task of doing the autopsies.

That is not what we are doing today. We are not digging our hands into gram-negatives of

dead people and then going and operating on someone. So, the conditions of Semmelweis were huge bacterial load, probably with gram-negative organisms. So, what Semmelweis showed was that washing your hands is a good thing. Semmelweis did not do a randomized trial of one product compared to handwashing alone or handwashing compared to nothing. We are not debating that Semmelweis was correct and that you need handwashing. What we are debating is handwashing with what, and how do we determine that that "what" is effective compared to just maybe plain soap and water? So, we are going to discuss further today what is known about surrogates in the setting of topical antiseptics, and Steve Osborne is going to go over this clinical correlation and tell us some more about it.

[Slide]

I would like to leave you with this quote by the statistician John Tukey which I think really relates to surrogates: Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which

can always be made precise. I will stop there. Thank you very much.

DR. WOOD: Thanks very much. It appears that we still don't have the slides from Michelle Pearson. Is John Boyce here? Yes? Good, so at least our next speaker is here. I suggest that we take a quick break right now and be back at ten o'clock and we will start again. We are hoping to get Michelle Pearson in before we do the questions. We will get back at ten o'clock.

[Brief recess]

DR. WOOD: Let's go to Dr. Boyce and then we will come back to Dr. Pearson, whose talk we do now have somewhere in the building, as they say, but we have been unable to play it yet. So, Dr. Boyce?

Antiseptic and Infection Control Practice

DR. BOYCE: Good morning. I am having some Power Point problems today because of a switch in versions so I hope this is going to work.

[Slide]

First I want to talk a little bit about

the importance of hand hygiene in preventing transmission of healthcare-associated infections. Most of you know that transmission of healthcare-associated pathogens often occurs via transiently contaminated hands of healthcare workers. For that reason, handwashing has been considered one of the most important infection control measures for preventing healthcare-associated infections. Despite this, the availability of published handwashing guidelines has not helped, and compliance with healthcare workers with recommended handwashing practices has remained low for decades.

[Slide]

This slide shows the percent compliance on the Y axis in 37 published observational studies of healthcare worker handwashing compliance. The main point here is that compliance rates varied from about 5 percent to 80 percent. The second point is that there is no trend towards improvement over this more than 20-year period. So, getting people to wash their hands as frequently as possible has

been a very difficult chore.

[Slide]

In 2002 the CDC published the guideline for hand hygiene in healthcare settings. I am going to briefly mention a few indications for hand hygiene that are listed. One is that it is recommended that we wash our hands with a non-antimicrobial soap or an antimicrobial soap if our hands are visibly contaminated with blood, body fluids or other proteinaceous materials. If the hands are not visibly soiled, then the guideline recommended the routine use of an alcohol-based hand rub for decontaminating hands in most other clinical situations. Alternatively, hands can be washed with an antimicrobial soap and water in other clinical situations.

The guideline recommends that healthcare workers decontaminate their hands before having direct contact with patients, donning sterile gloves to insert a central intravascular catheter, before inserting indwelling urinary catheters or peripheral IV catheters, and before eating.

[Slide]

It is recommended that we decontaminate our hands after having direct contact with a

patient's intact skin, like taking a blood pressure; contact with body fluids or wound dressings if our hands are not visibly soiled; after moving from a contaminated body site to a clean body site during an episode of patient care; after contact with inanimate objects in the immediate vicinity of the patient; and after removing gloves. So, there are a lot of indications for cleaning your hands.

[Slide]

In fact, the number of hand hygiene opportunities that healthcare workers have can vary considerably. In a large study, done by Dr. Pittet, they found that the average number of hand hygiene opportunities per hour of care was 24 in pediatric units, and the average was 43 per hour in intensive care units. In fact, the lack of sufficient time to actually perform this large number of handwashing episodes is a major factor

influencing poor handwashing compliance.

[Slide]

This slide shows the results of a number of observational studies where healthcare workers were observed to see how many times they actually cleaned their hands. You can see on your right that the average number of times per 8-hour shift was anywhere from 13 times to 26 times in an 8-hour shift. So, we are talking about frequent use of these products.

That sounds pretty frequent but let me present it another way, in a recent prospective trial that we conducted that involved 57 volunteer nurses working in intensive care units, a hematology-oncology ward and general medical ward, each nurse carried a portable counting device and prospectively clicked the counter every time they cleaned their hands. On the right you see a graph that, along the X axis, shows the number of hand hygiene episodes that these nurses recorded during a 3- to 3.5-week trial period. You can see that most nurses cleaned their hands anywhere from 100

to 450 times in a 3- to 3.5-week period.

[Slide]

So, one thing that is very clear is that, because of the high frequency of use of these products, providing healthcare workers with products that are well tolerated is very important. Poorly tolerated products result in poor compliance often because of irritant contact dermatitis, as shown in the picture, where this physician has bleeding knuckles after using soap and water handwashing 57 times over a period of a couple of weeks. Products that have a high degree of antimicrobial activity, that is, a high log reduction, but are poorly tolerated may actually be counterproductive.

[Slide]

Now, another important issue for which we have very little information is what level of log reduction of bacterial counts on the hands is actually necessary to prevent transmission of pathogens. As you know, the efficacy of these agents is often expressed as a number of log

reductions of bacterial counts on the hands of volunteers, 1, 2 or 3 log reductions for example.

Although the review of the literature that I did apparently is not as big as what FDA has actually done, I reviewed over about 700 articles and couldn't find any evidence regarding the number of log reductions that are necessary to prevent transmission of healthcare-associated pathogens. So, we just don't know how many log reductions we need.

[Slide]

Another thing for which I think there is little, if any, data relates to whether or not we need products that have a cumulative effect. As you know, the tentative final monograph requires that healthcare personnel handwash agents produce a 2-log reduction after the first wash and a 3-log reduction after the 10th wash, therefore showing a cumulative effect.

In the review of the literature that I did I failed to identify any data supporting the need for a cumulative effect. As a clinician with 25

years of experience working in hospitals, I am not aware of any evidence that patients who are cared for in the middle or at the end of a work shift are at higher risk of infection than those that are cared for at the beginning of a shift. I am also not aware of any evidence that patient care activities that are performed in the middle or near the end of a work shift result in greater hand contamination than those that are performed at the beginning of a shift. So, frankly, from the standpoint of a clinician or of infection control, I fail to see the logic in requiring a cumulative activity of this type of product given the way they are used and the types of patients that we take care of.

[Slide]

Another thing that actually has changed since the TFM was originally developed is the frequency of glove use. Since the late 1980's nurses, physicians and other healthcare workers use gloves far more frequently than they ever did in the past. A recent observational survey done of

nurses working on a general medical ward found that these nurses visited patients an average of about 54 times during an 8-hour shift, and they found that the use of gloves varied depending on the type of patient care activity. When the nurses were going to have contact with body fluids they wore gloves 86 percent of the time. If they were going to have skin contact only, then it was more like a little over 30 percent of the time that they wore gloves; even less frequently for equipment contact. So, in fact, glove use does vary among healthcare workers but it is certainly far more common than in the past.

[Slide]

A number of studies, shown here, have documented that gloves can and do reduce the level of hand contamination when they are worn. McFarland looked at hand contamination with C. difficile and found that 46 percent of healthcare workers who did not wear gloves contaminated their hands with C. dif.. No healthcare workers who wore gloves had C. dif. on their hands. Olsen and

colleagues found that gloves prevented hand contamination in 77 percent of instances. Dr. Pittet found that when no gloves were used and they measured hand contamination rates, they found out that the hands were contaminated with 16 CFUs/minute of patient care when no gloves were used, but only 3 CFUs/minute when gloves were used, showing the protective effect of gloves. Finally, Tenorio et al. found that gloves reduced the risk of hand contamination by vancomycin-resistant enterococci by 71 percent. So, in fact, to the extent that people do wear gloves during patient care nowadays, their hands are probably less heavily contaminated than they were back in the '60's, '70's and early '80's.

[Slide]

One thing that I thought that I was supposed to try to address was whether or not there is any evidence that the products that are currently on the market have any kind of clinical benefit in a healthcare setting. I wanted to mention this model by Ehrenkranz. It was a field

study that was supposed to reproduce clinical hand contamination. Nurses touched the skin of patients who were heavily contaminated with gram-negative bacteria. They then cleaned their hands. They either used plain soap and water handwashing or they used the 63 percent isopropyl alcohol hand rinse. After cleaning their hands, the nurses touched catheter material, like a Foley catheter type material, and then that catheter material was cultured on agar plats.

What they found is that bacteria were transferred from the hands of the nurses onto this catheter material in 11/12 experiments when plain soap was used to clean their hands but only 2/12 experiments when the alcohol hand rinse was used.

[Slide]

Now, in terms of clinical trials, which I think is a major issue as was discussed in part by the last speaker, this slide shows one sequential trial of three hand hygiene regimens. It was done in the surgical intensive care unit by a very experienced infection control physician. They

looked at non-medicated soap, 10 percent povidone-iodine or 4 percent chlorhexidine gluconate. Each product was used exclusively in the ICU for 6 weeks. Surveillance for nosocomial infections was performed. What they found was that the incidence of healthcare-associated infections was 50 percent lower during times when the two antiseptic-containing handwash agents were used, suggesting that these hand hygiene products that were available at that time reduced infections better than plain soap and water handwashing in this short trial which was only done in one ICU.

[Slide]

This slide discusses a prospective trial done to compare two hand hygiene regimens. It was a prospective trial with a multiple crossover design. It was done in three intensive care units in a university hospital that just happened to have one of the largest and most highly respected infection control programs in the country at that time. So, they had lots of resources relatively speaking. They followed over 1,800 adult patients

for nearly 8,000 patient-days at risk. The two regimens compared were 4 percent chlorhexidine gluconate versus a combination regimen of isopropyl alcohol and a non-medicated soap. Healthcare workers were told that when the alcohol and non-medicated soap were available they were supposed to use the alcohol routinely for cleaning their hands.

[Slide]

What they found was that the number of patients who developed a healthcare-associated infection was 96 in the chlorhexidine time period and 116 when the alcohol and plain soap were available. So, the incidence density was lower with the 4 percent chlorhexidine. The number of healthcare-associated infections was 152 during periods when the 4 percent chlorhexidine was used compared to 202 when the combination regimen was available--again, a lower rate with the 4 percent chlorhexidine. Infection rates were significantly lower in 2/3 ICUs when the chlorhexidine was used.

[Slide]

Despite this being planned by a very experienced and highly respected individual, with a large team working with him, this clinical trial

ran into some problems. First of all, the overall compliance of healthcare workers, as shown on the left, was not the same during the two trials. It was about 42 percent compliance when the chlorhexidine was available versus 38 percent when the other regimen was available in the units. The difference was actually statistically significant.

Another important problem that emerged, despite this trial being well planned and designed, was that the volume of the products used varied significantly. The amount of soap and isopropyl alcohol used when available was significantly lower than the volume of chlorhexidine used when that product was available. Even though healthcare workers were told they should use the isopropyl alcohol routinely when available, for reasons that are not either understood or discussed by the authors, the healthcare workers hardly ever used the alcohol. So, this trial was really more a

comparison of 4 percent chlorhexidine against plain soap and water for the most part.

So, one problem with this trial is that it is very difficult to control the activities of all these healthcare workers in all these ICUs over an 8-month period, and to get them all to do exactly the same thing and to do it with exactly the same frequency.

[Slide]

From the eyes of a beholder here who works in a hospital, that is one of the problems with clinical trials. When you use a nosocomial infection rate as the outcome measure for efficacy of hand hygiene agents, there are many, many confounding variables including host factors; the rate of importation of organisms from nursing homes or other sites into the hospital and onto the wards; the level of compliance of healthcare workers with recommended hand hygiene, with recommended barrier precautions, how frequently they follow guidelines for central line placement and for ventilator-associated pneumonia prevention.

If you are talking about surgical site infections you have to worry about the skill of the surgeon; whether or not prophylactic antibiotics were used and timed appropriately; and whether or not any active surveillance cultures are being done on the wards where the studies are being conducted.

So, from my viewpoint, there are so many confounding variables that that, in and of itself, makes the clinical trials extremely difficult to do and extremely costly. To me, it seems like the use of surrogate endpoints to assess efficacy of hand hygiene products still has merit.

[Slide]

I want to mention a little bit more about clinical benefit. None of the things I am going to mention are carefully controlled, prospective trials partly for all the reasons I have just mentioned. This one publication involved a surgeon whose hands, but not other body parts, were colonized with a virulent strain of Staphylococcus epidermidis that caused an outbreak of surgical site infections related to cardiac surgery. This

surgeon was using a noon-antimicrobial soap for a preoperative scrub because of previous problems with hand dermatitis so he followed the recommendation of his dermatologist.

An epidemiologic investigation that included case control studies and molecular typing clearly implicated the surgeon as the source of this outbreak, and we told him he had to stop doing cardiac surgery and to start using a 4 percent chlorhexidine gluconate surgical scrub. After he did so the outbreak terminated and we did not see that strain any further in cardiac surgery infections, demonstrating that the antimicrobial soap that was available didn't appear to have benefit.

[Slide]

An outbreak of vascular surgery-related surgical site infections occurred when an operating room was not provided standard povidone-iodine. The surgeons were used to using preoperative surgical hand scrubs. The vascular surgeons in the hospital decided to use plain soap for hand

scrubbing before surgery, while other surgeons used a 2 percent iodine with 70 percent alcohol for preoperative hand scrubbing. Hand scrubbing with plain soap was significantly associated with the occurrence of this outbreak of surgical site infections and reinstitution of povidone-iodine hand scrubbing terminated the outbreak, again suggesting that this povidone-iodine product had value in reducing surgical site infections.

[Slide]

Of course, the CDC guideline for hand hygiene was published in 2002 and the guideline recommends routine use of alcohol-based hand sanitizers for cleaning hands before and after patient contact as long as the hands are not visibly contaminated.

[Slide]

Not long after the guideline was published, actually in January of 2003, the Joint Commission on Accreditation of Healthcare Organizations sent out a sentinel event alert to hospitals and recommended that hospitals comply

with the CDC's new hand hygiene guideline. So, I think both the Joint Commission and CDC are standing behind the guideline.

[Slide]

This study was done where a 70 percent ethanol hand gel was introduced hospital-wide into the hospital. A multidisciplinary program to improve hand hygiene was carried out. During the following 12 months the alcohol hand product was used an estimated 440,000 times by healthcare workers and they found a consistent reduction in the proportion of all methicillin-resistant Staph. aureus that was hospital-acquired during the 12-month period.

[Slide]

This slide shows the impact of one of these alcohol hand sanitizers on the hand hygiene compliance in our hospital. Compliance rate is shown on the Y axis. Observational surveys conducted by the same infection control practitioners each time revealed that, by having this new alcohol hand gel available and promoting

its use and educating people about it, the overall hygiene compliance improved from 38 percent to 63 percent, and the proportion of all hand hygiene episodes which were performed using the alcohol hand gel, which is shown in the red part of the bars, increased significantly.

Not shown on this slide is the fact that the proportion of all methicillin-resistant Staph. aureus--let me put that another way, the proportion of all Staph. aureus isolates that are due to methicillin resistance in our hospital levelled off about the time that survey 2 was done, and actually decreased by 5 percent over the following year and a half. This decrease in MRSA in our hospital occurred during the same time frame when MRSA continued to increase in prevalence in the hospitals that participate in CDC's National Nosocomial Infection Surveillance program, or NNIS. Although it is rather crude data, we think that the hand hygiene program probably has helped reduced MRSA in our hospital as well.

[Slide]

In conclusion, conducting clinical trials to assess the efficacy of healthcare personnel handwash products is, in fact, extremely difficult,

expensive and, as far as I am concerned, largely not practical. If they are to be done, they are going to be very expensive.

Widespread experience with currently available products, combined with some of the epidemiologic studies that I mentioned, provide some evidence of their clinical benefit in healthcare settings. Multiple studies have shown that promoting the routine use of alcohol-based hand santizers, when combined with educational and motivational material, can improve hand hygiene practices among healthcare workers.

[Slide]

There are no published data that I am aware of demonstrating that cumulative activity of healthcare personnel handwash agents or surgical scrub products results in lower rates of healthcare-associated infections. Removal from the market of hand hygiene products that are currently

in widespread use in healthcare facilities would, in fact, disrupt national efforts to improve hand hygiene practices among healthcare workers. So, I personally would hope that there is no regulatory action that ends up removing a lot of the current products from the market because I am convinced, again on a personal level, that they do have value. Thank you.

DR. WOOD: We have received Dr. Pearson's slides from the wilds of Atlanta and we think we can show them. Is that right?

MS. JAIN: Yes.

DR. WOOD: Unfortunately, sort of like CNN breaking news, because the slides are just in we don't have a handout. We are going to have her on the phone. Dr. Pearson, can you hear us?

DR. PEARSON: I can.

DR. WOOD: As you go through the slides, Dr. Pearson, if you tell us when you want to change to the next slide, we will be able to do that. Let's go.

Prevention of Surgical Site Infections

DR. PEARSON: Good morning and thanks to the meeting organizers for tolerating my inconvenience and thank you for the opportunity to

present on the topic.

[Slide]

What I hope to do in the next few minutes is really to talk about some of the epidemiologic complexities of looking at the effectiveness of any preventive measure, whether it be cutaneous antiseptic or other preventive measures, using surgical site infections as the context for that discussion. Next slide.

What I am going to do is first provide an overview of what we know about the epidemiology of surgical site infections, including the incidence and risk factors for infection. I will talk next about some of the preventive strategies that have been shown to decrease that risk; highlight some of the current surveillance systems for monitoring the incidence of surgical site infections; and conclude with talking about how we, here at the CDC, go about developing our policies and recommendations

for prevention of healthcare-associated infections, such as SSIs. Next slide.

Just to give you a little bit of an idea of why this is an important topic and to frame it with some numbers, it is estimated that somewhere in the neighborhood of 20 million inpatient surgical procedures are done each year in the United States, and 2-5 percent of these procedures are complicated by a surgical site infection.

Based on our surveillance system, surgical site infection is the second most common healthcare-associated infection, comprising about a quarter of all of the infections reported to CDC. These infections come not only at a cost to the patient but also a cost to the healthcare delivery system. These infections result in anywhere from an additional week of hospital stay and they cost anywhere from $400 to $2,600 per infection, and these total well in excess, and approaching in some instances, close to a billion dollars a year in terms of healthcare dollars. Next slide.

In terms of the way we define or look at surgical site infections at CDC, we classify them either as incisional surgical site infections, and

those include superficial infections which involve the skin and the underlying subcutaneous tissue, or deep incisional surgical site infections which involve the underlying soft tissue as well. Obviously, the most severe and costly infections are those that involve the underlying organ or organ space surgical site infections and those involve really any part of the anatomy other than the incision that might have been opened or manipulated during the procedure. Next slide.

This is a cross-sectional schematic to illustrate just a little bit more clearly an abdominal wall that shows the various classifications. As you can see, a superficial incisional SSI would involve the skin and the subcutaneous tissue. A deep incisional SSI would extend down into the fascia and the muscle. The organ space surgical site infection, obviously, would include the organs in that surrounding

tissue. Next slide.

Now, when we look at the organ or the potential sources for the pathogens that result in a surgical site infection, overwhelmingly these arise from the patient's own endogenous flora. There are also secondary sources for the pathogens that result in a surgical site infection. Those can result from pathogens that are available in the operating room theater environment. They may result from operating room personnel that are in and around the surgical field or, not uncommonly, at the head of the table of the anesthesiologist. Less commonly, these infections may result from seeding of the operative site from a distant site of infection. Next slide.

If we look at the microbiology of the surgical site infections--and this slide is somewhat dated but suffice it to say that the distribution of these pathogens is still predominantly--the primary organism are staphylococcal infections, not surprisingly because these arise primarily from the patient's own

endogenous flora. The predominance of these pathogens is Staph. aureus, and then with certain procedures like cardiac surgery, and more recently we have been looking at some data from prosthetic joint infections, and it appears that staphylococci now account for in the neighborhood of around 50 percent of the infections causing surgical site infections. We have also seen an increase in the proportion of those staph. infections that are due to resistant organisms, such as methicillin-resistant Staph. aureus. Next slide.

Less commonly, SSIs may be due to some unusual pathogens, such as the ones shown on this slide that are typically due to either contaminated products or solutions that are used in and around the surgical site, or to colonized healthcare workers, again, that might be part of the surgical team. When you see clusters of infections that are due to these unusual pathogens you should think of a common source, such as the contaminated vehicle or potentially the colonized healthcare worker who is disseminating the organism. Next slide.

Regardless of where the organism arises, the pathogenesis of a surgical site infection can kind of be distilled into this numerical formula

and relationship shown here. That relationship really is a combination of the dose or the amount of bacterial contamination at the surgical site infection, the virulence of the colonizing or contaminating organism, and then the underlying sort of resistance of the host. Those three factors are really give rise to the risk of surgical site infection. Next slide.

If we look at some of the epidemiologic factors that have been associated with influencing the risk of acquiring a surgical site infection, they can be broadly categorized into those that are host- or patient-related factors, such as age, body mass index, obesity, the presence of diabetes and, as we will see later it may not just be a patient who is labeled with diabetes but having hyperglycemia at the time of surgery, the nutritional status of the patient, whether the patient has a prolonged preoperative stay, again,

whether there is infection at a remote site at the time of surgery, and whether the patient is on immunosuppressive medication such as steroids, or whether the patient is a smoker or uses nicotine.

Some of the procedural factors that have been associated with influencing the risk of surgical site infection are things like hair removal or shaving, the duration of the procedure, surgical technique, the presence of foreign bodies such as drains, and things like the appropriateness or inappropriateness of antimicrobial prophylaxis. Next slide.

What I am going to do now with the next series of slides is talk a little bit about some of these modifiable factors in terms of things that we recommend, or things that are recommended, to be done to minimize or moderate the risk of a patient acquiring a surgical site infection. Next slide.

There are a number of randomized, controlled trials showing the benefit of perioperative prophylaxis and I won't belabor you with those data. The feeling is that this is

probably one of the most important things that we can do in terms of modifying risk of infection. When we talk about antimicrobial prophylaxis we are really referring to a brief course, most commonly a single dose, of an antimicrobial agent that is given just before the operation begins. Antimicrobial prophylaxis is not intended as therapy. It really is a preventive strategy ,and it really should be used as an adjunctive preventive measure and not really used to supplant basic things like aseptic technique and some of the other basic principles of preventing surgical infection.

Now, antimicrobial prophylaxis, as I said, has been studied in a number of procedures, a number of well done randomized, controlled trials and it is shown that its use, if done appropriately, can decrease the risk of surgical site infection at least 5-fold. Next slide.

But surgical prophylaxis--again, to show you how complex this whole issue is, is not a matter of just giving an agent and giving the right

agent, but also giving it at an appropriate time. Now, this slide summarizes a study done by Classen, and I think it is one of the more classic studies looking at the importance of timing of antimicrobial prophylaxis in terms of its efficacy in preventing surgical site infection.

What Classen did was actually study nearly 3,000 elective clean and contaminated surgery. He looked at the timing of the antibiotic and its influence or relationship to the risk of infection. If you look at what he called early antimicrobial prophylaxis, that is antibiotics given 2-24 hours before incision, the rate of infection in that cohort was 3.8 percent. If he looked at antibiotics that were given postoperatively, that is 3-24 hours after incision, the rate of infection was 3.3 percent. If he looked at antibiotics that were given within 3 hours after the incision, the rate of infection was 1.4 percent. Lastly, the rate of infection was lower for antimicrobial prophylaxis that was given within 2 hours of the incision, 0.6 percent. So, again, it is not just a

matter of giving prophylaxis and giving the right agent, but this issue of timing is critically important. Next slide.

This next series of slides talks not only about this notion of giving antibiotics at a critical point before incision, but talks about the impact of prolonged surgical prophylaxis. This is a study that was a prospective study that looked at a cohort of CABG patients. They looked at those patients who received antibiotic prophylaxis within 48 hours of the procedure and those for whom the prophylaxis was continued for greater than 48 hours after the procedure. Next slide.

They looked at two outcomes, not only the incidence of surgical site infection but also the likelihood of acquiring a resistant organism if a surgical site infection did occur. Interestingly, what they found is that nearly half of the patients received antimicrobial prophylaxis greater than 48 hours after the procedure. Again, antimicrobial prophylaxis is intended to be given around the time of incision to get the maximal sterilization, if

you will, of the surgical site. But here we see that at least in half the cases patients are getting prophylaxis beyond two days after the surgery.

What they found is that the incidence of infection in this cohort of patients really was no different if antibiotic prophylaxis was discontinued within 48 hours or if it was continued for greater than 48 hours. But, interestingly, the rate of acquiring a resistant pathogen was 60 percent higher in those patients who received prolonged antimicrobial prophylaxis. So, again, antimicrobial prophylaxis and its influence on SSI is not only getting the right agent but getting it within the right interval and discontinuing it as soon as possible following the surgical procedure. Next slide.

Another area that I think is particularly intriguing as to the complexity of things that would have to be considered or controlled for in looking at SSI risk is this whole issue of glucose control and perioperative management of

hyperglycemia. This slide actually summarizes a prospective study that was done in a group of diabetic patients who were undergoing cardiac surgery, over nearly a decade at one hospital.

They had two groups of patients. Again, this is a prospective intervention trial with a pre- and post-design. The control patients were those who had received sort of the traditional therapy with their glucose being measured and monitored intermittently, and being given subcutaneous insulin. What they called the treated group were patients who were placed on a continuous IV insulin drip for the immediate operative period and for up to 48 hours postoperatively. Next slide.

The outcomes were that they looked at the levels of blood glucose that were below 200 mg/dL, and that was sort of the target level, within the first two days postoperatively. The other outcome obviously was the incidence of surgical site infection, and they focused on deep SSIs. What they found is that in the group who got traditional

management using subcutaneous insulin on a PRN basis the rate of surgical site infection was 2 percent as compared with the 0.8 percent in those patients who were managed with a continuing IV drip. This difference was highly statistically significant.

Now, there have been some subsequent studies that have looked at sort of the prevalence of patients who are hyperglycemic who don't carry the diagnosis or label of diabetes. Again, this notion of perioperative glucose management probably has broader implications beyond just the diabetic patient population. Next slide.

Another sort of titillating article that is summarized here and I think alludes to some of the complexity of this issue is this notion of perioperative oxygenation, the theory being that better oxygenated tissues are less likely to be at risk or be prone to developing an infection.

This was a study that was published in the New England Journal in 2000. It was a randomized, controlled, double-blind trial that looked at a

relatively small group, 500 patients who were undergoing colorectal surgery. Again, I want to emphasize that this was colorectal surgery. The intervention was that patients were randomized to receive either 30 percent or 80 percent inspired oxygen during and for up to 2 hours following the surgical procedure.

Now, what they found is that the incidence of surgical site infection was 5.2 percent in those who received higher 32 percent versus 11 percent in those who received 30 percent oxygen. That difference was statistical significant.

There has been a more recent study that came out in JAMA, and I did not summarize that here, looking at a more heterogeneous population of patients undergoing intra-abdominal procedures, again, randomizing them to receive 70 percent oxygen versus 30 percent inspired oxygen. That study concluded that there was not only no beneficial effect to a higher level of inspired oxygen but, in fact, there might be some detrimental consequences. In fact, they found a

higher rate of surgical site infections in those people who got more oxygen.

I say this to say again that this difference might be in part attributable to the population that was studied in terms of procedures. So, a lot of these things have to be factored in, in terms of trying to extrapolate findings from one cohort to another--not only what the intervention was but the population and the procedure that was studied. Next slide.

What about the issue of antisepsis and antiseptics? Probably, as you have heard from Dr. Boyce, a lot of the studies around the efficacy and the benefits of antiseptics really use bacterial count on scans and the amount of cutaneous flora remaining after their use as the primary outcome measure. When we look at hard outcomes or harder outcomes in terms of patient outcomes, data becomes much thinner.

These are just summarizing some data, and these are admittedly older studies and, you know, these studies to be done today are much more

difficult for a variety of reasons, but these three studies summarize data looking at surgical site infection rate with patients receiving preoperative showers versus those not getting showers. The earliest study was in the '70's where the rate among those who did not get showers was 2.3 percent versus 1.3 percent. In the subsequent two studies, in the 1980's, the actually the difference was quite closer.

Again, I think some of these studies, although they did not show a statistically significant difference, may be confounded by failure or inability to control for a lot of the factors that we mentioned up to this point. But, also, I am not convinced that these studies were adequately powered to detect a difference. Next slide.

Another factor that has been shown to influence the risk of surgical site infection is the whole issue of hair removal at the site of infection. In short, not unlike the story that I portrayed with antimicrobial prophylaxis, it is not

only a matter of do you remove hair or not remove hair but how you do it, and when you do it. They are all part of the complexity of influencing the risk of surgical site infection.

This is a study that, again, is admittedly old and I am not aware of this kind of study being done sort of in a more modern era, but if you look at those procedures where no hair removal was done, or hair removal was done using a depilatory, the rate of infection was less than 1 percent. In those procedures where a razor was used the rate of infection was nearly 8-9-fold higher in those first two categories of procedures. It is not surprising. Razors allow for microabrasions and nicks in the skin and, obviously, it is not difficult to imagine how those would be sort of easy portals of entry for any organisms that are left on the skin. Again, like I said, it is not only a matter of do you remove hair and how you do it but also the timing.

This study also looked at whether shaving done immediately prior to surgery, within 24 hours

of the procedure, or done later or much, much earlier, before 24 hours of the procedure--was that associated with a risk. As you can see, there was a nice step-wise progression with shaving or hair removal being done close to the procedure being associated again with the lowest risk. Again, one can imagine that that may be due to the immediate effect of skin cleansing. You have the benefit of perioperative prophylaxis being given in and around the time of the procedure. So, again, this is another issue that has multiple layers to it in terms of influencing the risk of surgical site infection. Next slide.

I will just say that the issue of clipping has been looked at in multiple studies, and it shows that, at least in terms of shaving, the clipping is associated with a lower risk or surgical site infection. Next slide.

I won't spend a lot of time on this but I just put this in to remind me to say that there are also data that suggest that the attire the surgical team wears, in terms of scrub suits or types of

suits, also may influence the amount of bacterial count in the operating room at the time of the procedure. I am not aware of any good data that link these type of things with hard outcomes like infection. Next slide.

I put this in to say that sort of the amorphous grab-bag term of surgical technique which, at least in epidemiologic studies, often manifests itself as a higher SSI risk being associated with a given surgeon is also something to consider, and actually it is fairly difficult to measure in an objective way. You know, it includes things like how they handle tissue; whether they eradicate dead space; whether they remove devitalized tissue; whether there are inadvertent things like entering a viscus; and obviously using things like foreign devices and leaving those in like drains and suture material. Again, these are all things that go under sort of a heading of surgical technique that are very, very difficult to measure in a systematic and objective way. Next slide.

I just want to say that although we believe the skin is the primary source of the pathogens that result in surgical site infection,

and most of our preventive measures are targeted at reducing that local contamination, there are things that are done in terms of the operating room environment to remove airborne bacteria that might also contaminate the surgical field.

I just put this up to show that the American Institute of Architects has established criteria for maintaining, if you will, the sterility or the ventilatory and environment parameters of the operating room. Those things include certain temperatures, relative humidity, air circulation and air exchanges. Next slide.

Just to follow on that, there are some data to suggest that air flow may have a role in SSI risk. This slide just shows some data, and again there are some issues with the studies and whether things were adequately controlled, and most of this data has been done with clean procedures,

particularly orthopedic procedures. This is a study that looked at 8,000 total hip and knee replacement. What they looked at was the role of ultra-clean air, laminar flow, antimicrobial prophylaxis alone or using those in combination.

What they found is that using laminar flow was associated with about a 50 percent reduction in surgical site infection risk among those patients undergoing total knee and hip replacement. Antimicrobial prophylaxis had a much larger benefit in reducing surgical site infection risk, going from 3.4 percent to 0.8 percent. When you coupled those, again, the additional benefit of laminar flow was not as marked compared with that of antimicrobial prophylaxis. So, again, part of these things are looking at the attributable fraction of any of these preventing strategies in terms of getting your bang for the buck. Next slide.

One thing that I have been asked by our colleagues at FDA is what does CDC monitor, and how does CDC track surgical site infections and many of

the things that happen in and around the time of operation. Next slide.

CDC has essentially three surveillance systems for monitoring healthcare-associated adverse events as they pertain to infection. The one that is really the component that is germane for this discussion is something called the National Nosocomial Infection Surveillance system, of the NNIS system. The NNIS system has been around for 30 years. It started in 1970. It measures nosocomial infections in patients who are critically ill, primarily ICU patients. It also measures infection in surgical patients. Next slide.

If we look at the characteristics of the hospitals participating in the NNIS system, the NNIS system is comprised of about 300 hospitals. There are roughly 5,000 to 6,000 hospitals in the United States so the NNIS system is comprised of less than 10 percent of the hospitals in the United States. These hospitals tend to be largely large academic teaching institutions. Nearly 60 percent

of them are teaching hospitals. The remaining group of hospitals has some sort of teaching affiliation. The hospitals in the NNIS system have a median bed size of around 360 beds, and there are no facilities in the NNIS system less than 100 beds. That is important because 50 percent of the hospitals in the United States are less than 100 beds. So, whether the data we see collected in the NNIS system are representative of all hospitals I think is one thing to consider. Next slide.

When we look at the specific data and variables that are collected in the NNIS system as they pertain to surgical patients, there is some basic demographic information like patient age and gender, their ASA score which is a measure that the anesthesiology colleagues use for sort of measuring the severity of illness of patients. They collect data on wound class; whether the operative site or the surgical site is related to trauma or not; the type of anesthesia; whether the procedure is emergency or elective procedure; the duration of the procedure; the length of postoperative stay;

the infection site; the infections pathogen. Is there any SSI-related mortality, as well as hospital demographics. Importantly, this system does not collect data on many of the processes that we have talked about in terms of influencing the risk of surgical site infection. Next slide.

One of the things that the system does is that it generates rates that can be used as national benchmarks for institutions to essentially measure their performance based on a given procedure, for example CABG or what-not. I think you have in your handout the most recent NNIS report that shows the national benchmarks for various procedures. An important part of coming up with those numbers is this notion of risk assessment. Part of that adjusting procedure is looking at something that is called the NNIS risk index. Again, that risk index is the composite score of the American Society for Anesthesiologists, or ASA, score, the wound class at the time of surgery and the duration of the procedure. These are the three variables, at least

that have been studied in the NNIS system, that have been shown to be most predictive of a patient's risk of developing a surgical site infection. Next slide.

These are some temporal trends in what we have observed in terms of surgical site infection rate over a period of the late 1980's to approximately 2000. Essentially, this is stratified by those patients who have procedures that are low, medium low, medium high and high risk. What you can see is that the lowest risk procedures in patients the rate of surgical site infections is actually quite low and has remained quite low. There has been a slight downward decrease in the middle categories, and again some of those rates are relatively low. But impressively, there has been a marked decline in the rate of surgical site infection among the highest risk procedures and patients. Again, you know, one question you might have is can you superimpose on this, or do you know how some of these various preventive strategies relate to this

graph, and we don't have procedure and patient specific data on who got prophylaxis at the right time, for example, and the risk of infection. Next slide.

I think an important thing in terms of this notion of designing any study or measuring the effect of any intervention is this notion of having good surveillance data or good capture of patients who undergo these procedures. In the NNIS system all of the patients who are enrolled in the system and recorded in the system are followed for at least 30 days postoperatively to monitor for risk of infection. If the procedure involved an implant such as a prosthetic joint the period of surveillance is up to one year for the risk of infection. These are very, very long periods of time of follow-up, and I think if you look at many studies the patients may not all have been followed for this length of time. Next slide.

Having said that, following patients for this period of time to meet this definition, it really has become more complicated if you look at

some of the trends of what is happening with healthcare delivery in the United States. I will focus your attention on length of stay, which has decreased at least by a third--and this was based on 1995 numbers; it is probably even lower now--and also look at the number of procedures that are actually being done in patients those have decreased, again based on 1995 data, by 25 percent. So, the ability to capture these patients requires a lot more effort and energy if they are going to be followed for 30 days postop or, in the case of an implant, up to one year postoperatively. In fact, our data would suggest that somewhere around 20 percent or less of the procedures that are complicated by an SSI is that surgical site infection detected during the admission where the procedure was done. Obviously, if the patient is readmitted because of some organ space infection we would capture those, but for lesser and some of the higher volume procedures that are primarily superficial infections, those people would never come back to the hospital. So, you have to rely on

a strong system of post-discharge surveillance to capture any untoward event and minor untoward event such as a surgical site infection. Next slide.

We, at CDC, are actually undergoing a transition in terms of our surveillance activity. I alluded to on the other slide that we sort of have three components to our surveillance. We have a dialysis surveillance network. We have something called NaSH, which is the National surveillance system for healthcare workers, and then we have the additional NNIS where the focus is on patient outcome. Those are all being rolled into one system called the National Healthcare Safety Network. Next slide.

NHSN, although it has a new name and it is a hybrid of all of our surveillance systems, maintains the same goals of the predecessor systems. The reason for doing this is that NHSN is going to be a web-based application which we believe will minimize a lot of the data collection burden and mangled data entry that the current system has. We are hoping that this system will

also increase the capability to capture electronic data, whether it be from laboratory information systems, administrative data bases, operating room records which capture a lot of the process things around the surgical patients, as well as pharmacy data to look at things around prescribing. Next slide.

Importantly, one of the priority areas for the National Healthcare Safety Network is really this notion of including process measures. These process measures will allow you to link them to outcomes so, for example, we will be looking at surgical prophylaxis as the first cut and whether the patient got the appropriate antibiotic based on national guidelines for that procedure; whether they got the antibiotic within a certain time, in this instance within an hour before the incision; and whether antibiotics were discontinued within 24 hours of the procedure. That will be able to be linked to outcomes data on patients. So, we will have some measure of how process relates to outcome. Next slide.

The last thing I will talk briefly about--and I was asked by FDA colleagues to give you a little bit of a glimpse of how we here, at

CDC, go about developing policy around some of these preventive strategies. Next slide.

We here also have a federally charted advisory committee, the Healthcare Infection Control Practice Advisory Committee, whose mission is really to advise the Secretary of Health and CDC about issues related to the prevention and the surveillance of healthcare-associated infection and related adverse events such as antimicrobial resistance in healthcare settings. Next slide.

The charge of the committee's activities and recommendations are really targeted and aimed toward clinicians, infection control professionals, regulators, purchasers and public health officials. The target setting for these guidelines--they were traditionally geared toward procedures and practices that occur in acute care settings but now these guidelines are really aimed to address procedures and healthcare delivery across the

continuum, including outpatient settings, home care and long-term care. Next slide.

These recommendations are aimed to be evidence-based, and all of the HICPAC guidelines are ranked. The recommendations are ranked to show the strength of the evidence. I won't read through the definitions of the categories; you can do that. But essentially there are three broad categories. The category I recommendations are in large part based on evidence or well-designed experimental studies or epidemiologic studies; the category II recommendations where there may be some suggestive evidence but this category may be based on expert opinions; and then the last category is for those practices for which there is either insufficient evidence or a lack of consensus regarding efficacy, in which case the committee would consider that practice or that recommendation an unresolved issue. Next slide.

What this does is actually sort of summarizes the categorization scheme and what it means regarding evidence and recommended practice.

In short, the difference between I-A and I-B is really the strength of the evidence but, in short, category I recommendations are those practices for which there is strong evidence supporting it and the implementation of that practice essentially is recommended for all hospitals. Category I-C--we added this fairly recently--are those things for which there might be legislation or federal or state mandates, such as the blood-borne pathogens standards for example, that says that all hospitals have to do this. There may or may not be good evidence supporting this but, because it is required by regulation, all hospitals must do it.

The category II recommendations, again, are those practices for which there is good or some evidence that the practice may be beneficial and that practice is suggested for implementation in many, if not all, hospitals. Lastly, the category of no recommendation are those practices for which there is insufficient or contradictory efficacy, that is to say, you might have four studies of equal quality, two showing a benefit and two

showing no benefit, in which case the recommendation for implementing that practice is an unresolved issue. Next slide.

Now, we too, as I am sure you advisory committee and many other advisory committees, have many challenges in trying to take this evidence-based approach to developing our policies. Sometimes we identify subject matter experts who are not necessarily methodologic experts in terms of conducting systematic reviews. Systematic reviews are labor intensive and costly so we often have resource limitations for doing that.

In our field of infection prevention and infection control, we don't have a body of randomized, controlled trials that, say, might be in the cardiology literature or some of the other more clinically based specialties so sometimes we have to rely on observational studies, which in many instances, by some, are considered a lower quality of evidence.

Lastly, our user needs, not uncommonly, outstrip what the available science there is to

support or to provide evidence-based recommendations. This is particularly true when we look at non-hospital based healthcare settings. Next slide.

Just to say that our guidelines come in three parts. The first part really is a comprehensive synthesis of the literature review and the research that establishes the scientific rationale for the recommendations that are contained in part two. Part two are the summary of the practice recommendations with categorization. More recently, we have now added a third part which outlines or provides three to five what we call performance indicators or performance measures that institutions can use to monitor their success in implementing these guidelines. These three to five indicators are category I-A recommendations, those recommendations or those practices that we believe the data suggest have the strongest impact on reducing that outcome. Next slide.

To conclude, what I hope I have done is show you that some of the complexities involved in

surgical site infection prevention are some of the things that have to be considered in designing any study to look at the effectiveness of any one strategy. This prevention really is a multifaceted approach targeting pre-, intra- and postoperative factors.

Our current surveillance systems really are limited in that they don't collect data on perioperative processes. Another thing I think complicating it that would have to be factored into any study to look at surgical site infections and impact of any measure would have to consider the fact that we have experienced a fairly dramatic shift in where surgical procedures are occurring, and that patients are staying in the hospital for a much shorter period of time. There would have to be some system in place to capture events that occur post-discharge or for procedures that are done outside the traditional acute care setting.

I will also say that in general the incidence of surgical site infections, in large due to advances in preventive strategies, is low. So,

studies that would look at any intervention would likely have to have a fairly large sample size.

Finally, some of the prevention practices, such as hand hygiene, might be very, very difficult to study using the traditional randomized, controlled, research design because you wouldn't randomized someone to do it or not to do it.

I will just conclude by saying that prevention is, obviously, primary, one of our primary focuses here, in our division, and many of the things that I have talked about specifically as guidelines, HICPAC guidelines, are available on the web site and that URL is in your handout. I think I will stop there and let you ask any questions. Thank you.

Question and Answer Period

DR. WOOD: Thank you. I guess what we will do is keep you on the line. I am told it will be technically difficult to do that once we start questions for other speakers so perhaps we could have the committee focus first just on Dr. Pearson, with questions for her.

Did I understand correctly that none of your surveillance instruments use outpatient surgical centers? Is that right?

DR. PEARSON: You are right. The current NNIS system does not. The NHSN, which should be going live in a few months--what it is going to now do is allow any facility that, for example, does surgery to report to the system. If you are an ambulatory surgery center you can also report your data to the system.

DR. WOOD: But even the large hospitals that are in the system right now that have outpatient surgery facilities, where these patients are not admitted, would not be in the system. Right?

DR. PEARSON: That is correct.

DR. WOOD: All right. Any questions from the committee? Yes, Mike?

DR. ALFANO: Thank you, Dr. Pearson. That was a wonderful presentation. I have a question about how to potentially explain the increase in nosocomial infections per 1,000 patient-days. As I

think about your database, it was occurring as managed care was coming in and, obviously, patient days were getting shorter per procedure. So, I wonder how much the increase per 1,000 patient-days relates to the difference in numbers of patient-days per se, which are going down so that someone, you know, could have acquired an infection at a comparable rate but the numbers would make it appear to be somewhat higher.

Also, a point that I think the Chair was getting at, there are more outpatient procedures and I think the tendency is that healthier patients are done in an outpatient setting which means they would be less likely to be candidates for infection. Could you project how much of the increase could be related to those types of changes in the inherent system as opposed to actual problems in hospital-acquired infections?

DR. PEARSON: Yes, let me just challenge a little bit your initial assertion that they are increasing. We are actually looking at some updated numbers. I think most of you are aware of

the number two million infections and the like, and what we have actually seen is that the actual overall number has gone down over the last decade or so. I think it is 1.7 or something.

But you are right, what we certainly have seen and believe is that the people who are in hospitals or getting inpatient procedures are sicker than they were a decade ago. So, you have a population at higher risk for infection so that certainly plays into the rate that we see.

You are right, consequently the lower risk patients are sort of skimmed off and are not getting reflected in these numbers that we are seeing, but also the people that are actually getting into the hospital and getting inpatient procedures are older, sicker, and have many more co-morbidities than one would have seen before; the 20 year-old is not being hospitalized now. Does that answer your question?

DR. ALFANO: Yes, thank you.

DR. WOOD: Yes, Jan?

DR. PATTERSON: Michelle, this is Jan

Patterson. Could you elaborate on what the CDC guidelines say regarding the surgical prep chlorhexidine versus alcohol versus betadine? As I recall, there is some discussion about the superiority of chlorhexidine used as an antiseptic but there is no specific recommendation of one over the other.

DR. PEARSON: Yes, that is right. The current guideline actually looks at a variety and does not recommend one specific product over the other in terms of surgical site prevention. In a more recent guideline around prevention of IV catheter-related infection we did specifically recommend chlorhexidine as the preferred agent for cutaneous antisepsis. Povidone-iodine can be used as an alternative but we did recommend chlorhexidine preferentially, in large part because there are now several randomized, controlled trials and even a meta-analysis which shows that chlorhexidine was superior to povidone-iodine in preventing catheter-related bloodstream infection. I think similar rigor, at least to my knowledge, in

terms of those kinds of head-to-head comparisons for prevention of surgical site infection is not available.

DR. WOOD: In the absence of any other questions for you, can you stay on the line? I guess the sound person can hear you so if you can hear us you can respond to that if you want. Will that work?

DR. PEARSON: Yes.

DR. WOOD: All right. Questions for the other speakers then? Yes, Dr. Larson?

DR. LARSON: Thank you. I would like to describe what I think is the current cyclical scenario that we are in right now that may explain why it is that there is very little evidence, and I totally agree with that, of a link between log reduction, how much we need in infection and also whether the TFM recommended procedures are the right ones that we should do.

I have been doing funded research on skin antisepsis since the late 1970's, right after the first TFM came out. I learned in my first couple

of studies that the healthcare personnel handwash recommended protocol testing in the TFM did not work for what I wanted to do clinically for several reasons. First of all, it is very difficult to reproduce. I learned that in various hands you can change the results you get simply by changing the amount of time that you allow to dry--just little, tiny changes in the protocol can change hugely the results you get. That was concerning although I know that the labs that do it, do it very well but there is a lot of room for variability in the test.

Secondly, we learned early on that by putting Serratia marcescens on the hands we could not decontaminate the hands after they were contaminated, and we found Serratia on our subjects' hands as far away as six days after putting it on. And, we felt it was unsafe.

Thirdly, by using paid volunteers, it really had very little to do with what is going on in field studies, etc. So, I stopped using the healthcare personnel handwash protocol in the lab setting because it simply wasn't clinically very

relevant.

So, what happens then is you have three groups that can possibly fund these studies. There is industry or there is NIH, or whatever. Industry can't really do studies with clinical endpoints because they need to link up then with somebody who is in a clinical setting. The labs that are doing the testing, are doing it very well in humans but not with patients, etc. They can't do studies on their own with clinical endpoints unless they link with somebody in the clinical setting. So, that leaves the researchers in clinical settings, like me, like John, etc. Then we need to get funding. We are in academic settings and, you know, we can get funding from industry but the price of the studies is prohibitive often and there is not a lot of incentive to look at clinical endpoints sometimes.

In the last three years I have been the PI on three NIH-funded grants to look at skin antisepsis. Each of those grants costs over a million dollars. One of them is already published,

and that was a study in the home setting so it is not relevant here. That was published in The Annals of Internal Medicine. The second one, which was a study comparing alcohol and CHG in neonatal intensive care units will be coming out in a couple of months in The Archives of Pediatric and Adolescent Medicine. The third one, which is funded again for over a million dollars, is a study to try to assess the impact of the new CDC hand hygiene guideline on infection rates in 40 hospitals. However, this is not assessing efficacy; this is assessing effectiveness.

So, one of the things we need to be clear about is what is FDA's interest. Are we interested in assessing efficacy or effectiveness? There is never going to be a clinical study that is going to look at efficacy because of all of the confounding factors, and I will be the first to admit that every study I have done has a lot of problems because there are confounders, etc., etc.

Judging from that, I think in some ways--because we have been dealing with this issue

since 1978 and I have been at several of these over the last decades--in some ways the horse is out of the barn. Now the Joint Commission has said to hospitals to get accredited you have to use the hand hygiene guideline. Therefore, it is not possible to get permission in clinical settings to do studies where you are comparing plain soap and an antiseptic soap because the hospital will not get accredited. So, it is too late in some ways.

Now, I think what has happened is that short-term political will has ended up, as it sometimes does with decisions to not fund the ideal study--you know, 20 years ago or whatever, if it were possible to do--has resulted in spending more money and time than we should have. So, I think that the published studies will never answer the efficacy questions in the clinical studies that need to be done.

My feeling is that our position right now for this committee is two choices. NIH doesn't want to keep funding these studies; they are too expensive. So, either FDA defines an ideal

protocol and helps fund the study--and I know you are not a funding agency--because nobody else will do it, or we just decide that we are going to look at safety and efficacy and if a product meets a certain standard, then we keep it on the market. But to look at clinical effectiveness, you know, unless the FDA is going to chip in with a little bit of money, NIH is not going to keep funding these studies.

DR. WOOD: Well, I am a lot more optimistic than that. I am not saying that is what we should do but if, for example, we recommended that efficacy studies were required you would find that industry would get them done in a heart beat. That has been my experience in the past.

DR. LARSON: Industry is doing the efficacy studies--

DR. WOOD: No, I am talking about efficacy in terms of clinical endpoints. There is certainly, you know, plenty of experience doing extraordinarily complex trials by industry funding that have resulted in clear demonstration of

efficacy or not. And, all of these trials cost huge amounts of money, certainly many times the numbers you are talking about. Any other questions? Yes, Frank?

DR. DAVIDOFF: I was curious how the initial or the existing recommended log reduction numbers were chosen because it seems pretty clear that they were, in a sense, pulled out of thin air. That is to say, there wasn't good, hard evidence on which to base them certainly in terms of clinical endpoints. So, there must have been some logic as to choosing the 1-log, 2-log, 3-log reductions as the specific numbers or in a sense threshold numbers or qualifying numbers to use as the criteria for judging these products. So, that is part (a) of the question.

The second, related part is why reductions were chosen rather than some absolute threshold number, rather than a relative number like a change. It seems, sort of from a biological standpoint or clinical standpoint, that it is not so much whether you have dropped from a million to

100,000 bugs but the more important point might be to get yourself below 100 or some other absolute threshold.

I was curious how those decisions were made because, if those are the ones we are going to stick with, it would be nice to know that there was at least some reasonably compelling logic behind those initial decisions.

DR. WOOD: Well, my reading of the briefing book was that there was not, but does somebody want to add to that?

DR. LUMPKINS: Yes, I will take a stab. Basically, the effectiveness criteria evolved based on our experience with the evaluation of NDA data. Basically, our effectiveness criteria are based on our experience with the performance of chlorhexidine gluconate in studies very similar to the ones that are in the TFM at this point.

DR. WOOD: But I think what Frank is asking is, as I understand the briefing document, you sort of saw what you saw for chlorhexidine and you used that as a kind of standard moving forward.

DR. LUMPKINS: Right.

DR. WOOD: And what he is asking is was there any data to link that to a clinical outcome.

DR. LUMPKINS: No.

DR. WOOD: Right. Then the second question he was asking was are there any data that relate absolute numbers of colony counts, or something, that would--

DR. LUMPKINS: The unfortunate situation is that the virulence of these organisms varies. So, you can pick one but we don't really have a good handle for most organisms so you would be forced into a situation where you would pick one organism arbitrarily which may or may not tell you something about the general population.

DR. WOOD: Okay. Tom, did you have a question?

DR. FLEMING: I do, and I would like to pose it in the context of John Powers' slide number 36. So, if we could take a moment to get that?

DR. WOOD: We will work on getting that slide up. In the meantime, Mary?

DR. TINETTI: Two quick questions. One, are there other examples like this where FDA has a standard for a surrogate that has never been linked to an outcome? Because the other examples that you had in your slides, John, were all surrogates that were linked to a clinical outcome.

Number two, these are all log reductions. Do we have any data on individual people, percentage of people who respond and don't respond to these?

DR. POWERS: I think what we usually try to do and what I tried to put in those slides as far as timing is that today, in our current regulatory environment, we would try not to do that where there was no link. What we like to do for serious and life-threatening diseases, like for HIV, is propose a plausible link and then study it. In HIV there were actually over 5,000 patients in which that viral load was validated. Actually, we had an advisory committee on that back in the late 1990's. So, it is important to realize that what I put up there is that this was developed in the

1970's before any of our current regulatory strategies.

DR. TINETTI: I understand that but are there any other examples? Is this the only example?

DR. POWERS: Not that I can think off the top of my head, no. Even if there was, it is not an example we want to replicate.

DR. WOOD: Yes, Dr. D'Agostino?

DR. D'AGOSTINO: Thank you. With regard to asking some questions about the design, could you say once again why the multiple wash is done in some of the studies? Because the industry is suggesting dropping it and there must be something more compelling about that than that it was just historically done.

DR. LUMPKINS: Unfortunately, a lot about the design is lost to time and I am not well versed in it. I can tell you what I believe to be the case. These are multiple use products. These studies were intended to simulate the actual use of the products. I almost feel like they were trying

to get more than one piece of information from these studies, one of them being the effectiveness over time and the other one being the potential for irritation.

DR. D'AGOSTINO: In the studies you were looking at the log reduction. We don't have an irritation measure that comes out.

DR. LUMPKINS: No, we absolutely don't but sponsors do routinely gather that information from those kinds of studies. If you look at the published literature--

DR. D'AGOSTINO: No, I understand that. I am just trying to figure out why we see it in the recommended designs. Thanks.

DR. WOOD: Dr. Taylor?

DR. TAYLOR: I would like to thank the presenters for their thorough presentations. They were quite useful to me because after I read most of the big book I was a bit more confused than I was before I started it. I still am to some degree. I think in the initial presentation that Dr. Susan Johnson made, in slide 10 she pointed out

that the current decision thresholds are based on NDA performance. There decisions regarding these agents are very complex, as Dr. Powers so eloquently pointed out. In Dr. Johnson's presentation, she said any change should be data driven.

I think if you are going to use that as your threshold for changes, we are in deep trouble because I think clinical outcomes versus these outcomes in these trials are quite different and it is just a complex situation of a moving target. So, I just bring that up as a point of beginning the discussion. I guess my optimism is not that high that we could actually help you with changes unless they were very specific things that you wanted to change.

DR. WOOD: If you could get the slide up for Dr. Fleming? Tom?

DR. FLEMING: I would like to just expand slightly on Dr. Powers' eloquent presentation. One of the very important observations is that when you are looking at biomarkers, for example here, it is

very important to understand whether, for example, lower levels of bacteria are associated with lower levels of infection. But it is critical, as should be clear from this presentation, that that just gets your foot in the door. That doesn't begin to validate the biomarker and it is entirely possible, if not highly likely, that you could then induce reductions in bacteria and not, in fact, reduce inductions in the infection rate. In fact, the correlation that exists there might not even lead you to be able to conclude that it is a causal pathway. I think that is expanding a big on what Dr. Powers was pointing out.

A simple example of this in infectious disease is mother to child transmission of HIV. We know that a mother that has a higher level of viral load has a greater risk of transmitting HIV to her infant. We know the higher the level of the viral load, the lower her CD4 count. So, we have strong correlations between the mother's CD4 count and her risk of transmitting HIV, and you can intervene with that mother in the month before labor and

delivery and you can give IL2 and that is going to spike her CD4 and it is going to do nothing to alter the risk of transmission of HIV because it is not the causal mechanism by which transmission is occurring even though CD4 is highly correlated.

In essence, what we need in order to be able to validate surrogates is precisely on this slide. You need both columns. You need trials that establish both the effect of the intervention on the biomarker, in this case log reductions in bacteria, and the corresponding reduction in rates of infection.

Dr. Powers gave a success example of cholesterol lowering but it is important to drill down on that success example. Gordon did a meta-analysis of 50 trials looking at fibrates and vitamins and diets and showed that it was an inappropriate surrogate because we were looking at 10 percent reductions in cholesterol that didn't predict an effect on MI or death. Statins came along with 40 percent reductions and we did see benefit, although as Dr. Davidoff pointed out, some

statins actually might have other mechanisms as well.

So, the message here is we need an array of trials that look simultaneously at what the level of effect is on the biomarker and what the level of effect is on the clinical endpoint. If cholesterol lowering is any hint of what might happen, lower levels of effects on the biomarkers, maybe a 1-log reduction won't translate into benefit where higher will. That remains to be seen but there are precedents for that type of phenomenon and we are only going to understand it when we follow this slide and we have studies that look at both.

DR. WOOD: Right, and just to add to that and sort of supplement what Dr. Larson was saying, we are spending as a country billions of dollars on the implementation of these strategies without knowing whether they work. So, justifying spending the money to find out whether they work seems to me a relatively trivial issue. Jan?

DR. PATTERSON: You know, talking about

the CD4 count and the viral load and, you know, the CD4 count not being predictive of the outcome there, I think it is also an over-simplification to say that antisepsis that is a clinical endpoint in decrease of infections in patients because the most common infections that we see and monitor are things like surgical site infections, bloodstream infections, ventilator-associated pneumonias that we know have multiple other factors that are probably more important, like the devices themselves and all those surgical factors that Dr. Pearson reviewed. But we also know that because of the mode of transmission of some diseases that can be transmitted in the hospital--conjunctivitis, for instance, which we know can spread like wildfire and can be fatal for immunocompromised patients, we know that is because people who have it rub their eyes and then touch patients and touch each other; and influenza which we know, and we are seeing this year, can go between patients and healthcare workers in a hospital, and multi-drug resistant pathogens, C. dif., all those things--you know, I

think that antisepsis question is more pathogen specific than all these device-related infections that we typically monitor. So, I think that saying that the clinical endpoint of infections overall in patients is a bit of an over-simplification for looking at antisepsis itself.

DR. WOOD: But isn't that also true in every disease? If we pick the example of cholesterol, heart attacks are not just due to cholesterol elevation; they are due to activation of endothelial factors, platelet activation, and so on and so on, all of which eventually summate to an MI but cholesterol is one risk factor. So, it seems to me to be true here. We are sort of discussing this as though this is fundamentally different from every other issue and I am not persuaded personally that it is.

DR. PATTERSON: Well, I think that the device aspect of it--I mean, we know that, for instance, from bloodstream infections and ventilator-associated pneumonias, ventilator-associated pneumonias in particular, the

device itself is the major risk factor; the same thing for urinary catheter infections, but the infection may be more likely be due to a patient's own flora rather than, say, a multi-drug resistant organism that is going around if antisepsis is in place. So, I think, you know, if we are looking at the big picture overall of infections it is a little bit difficult to apply that specifically to antisepsis.

DR. WOOD: Doesn't that speak to drilling down more to the infections? For instance, if you are going to a strategy to prevent eye to patient transmission you would have a specific strategy for that. Surgical site infections would be something different. Ventilator infections would be something different, like, you know, HIV versus cholesterol reductions or whatever.

DR. PATTERSON: Well, I think that is one of the difficulties we have been discussing because in every outbreak investigation intervention we don't just do a single factor; we do multiple things.

DR. WOOD: Right. Dr. Powers I think wants to respond.

DR. POWERS: I wanted to get to what Dr.

Larson said and reiterate this question too. One of the things when I showed some of the things that may impact from an intervention going to the clinical outcome, down at the bottom was other factors that affect that clinical outcome. If it turns out that those other factors--and Dr. Pearson enumerated a number of them--are far more important than what we are doing with antisepsis, that answers the question of effectiveness which, in this setting, is the paramount one. It doesn't matter if we get rid of the organisms if doing that has minimal impact on those other mechanisms of disease which actually result in the actual clinical outcome. So, saying that we are doing something--it is circular reasoning, saying doing something must be effective because we changed the organisms but all those other confounders makes it look like it is not. So, I think the effectiveness question here, as Dr. Larson said, is very

important.

The other thing I wanted to get to is the JCAHO question, having learned all this myself in a regulatory agency. The Center for Medicare and Medicare Services contracts to organizations like JCAHO to accredit hospitals. JCAHO does not stand by itself and does not make those regulations. We have actually worked very closely in certain situations with CMS, and they are very interested in this issue of do these products work or not because, as Dr. Wood said, there is an awful lot of money getting spent in this situation. So, we have worked with them in other situations, and we have not discussed this particular one with them in terms of how do we get this information that we need in order to be able to know whether what we are doing is actually effective.

DR. WOOD: Right. Dr. Leggett?

DR. BRADLEY: Two comments, one to elaborate on what John and Tom have been saying with respect to trial design and the strength of the evidence, certainly over the past five to ten

years how anti-infective drugs that are administered systemically are evaluated has become much more stringent based on animal models, based on mathematical modeling, in vitro characteristics of all these anti-infectives on organisms, the ability of drugs to get to the tissue sites--all sorts of things. It seems as though the evolution of this particular field began in the '70's when we had far fewer tools by which to evaluate things.

In looking through the 1994 Federal Register document, there were some references to the issues raised by Frank regarding what the inoculum needs to be to cause an infection, and I saw some reference to a 1950 article in which a study was done where volunteers received injections of staphylococci into the skin to see how much you need to put in. I don't think you could get that past our human research committee these days but animal model studies are now what we use in that context. I couldn't find the animal models within those several hundred pages. There was something on shaved rabbits with iodinated iodine and shaved

primate backs, but nothing that you would expect where there was a surgical animal model which I think would be very helpful. Even though animals aren't humans, it would be a first would be a first step. So, the question is are there any of those studies that were ever done in animal models that could help us begin the process?

Secondly, there was some ambiguity on cumulative effect of these topical antiseptics. From the presentation that Michelle Jackson made earlier, on slides 12, 13 and 14 there is a one-day cumulative effect for healthcare personnel handwashes where, as I understood it, during one day there are ten handwashes and they are sampling at the end of that tenth handwash which shows a three-log reduction. That is in contrast to the surgical hand scrub cumulative effect in which there is a five-day cumulative effect sample. When people say cumulative effect, those are two huge differences to me and I am not sure which one we are talking about.

DR. WOOD: Well, these are the proposed

reductions rather than clinical trial demonstrated effects. Right?

DR. BRADLEY: Industry was saying that one of them was in error and one of them was correct.

DR. WOOD: The real Dr. Leggett?

DR. LEGGETT: Thank you. First let me respond to John. Yes, there are animal models for surgical site infections. I know the Vanderbilt group has also looked at that in the context of Staphylococcus aureus producing beta-lactamases that tend to degrade ceftezole more than others, and there are mouse models of skin and soft tissue infections, and that was going to be one of my points too.

The other thing is in terms of other animal models, dogs and prophylaxis, when we talk about timing and tissue levels preceding our use of ceftezole, you know, in a wide context for surgical site infections. So, I think that with a little digging we can find those things.

I wanted to go back to John's slide number 36 again in the context of what Tom talked about,

trying to correlate clinical endpoints and surrogate endpoints and use of neutralizers in the studies. If we are neutralizing clinicians' hands, why are we neutralizing for the gloves? If there is a difference between chlorhexidine and soap and water, the study that was just passed to us last week showed that, quote, reduction of CFU is the same for just soap and water as it was for all the other products, something doesn't jive there. So, maybe one of the rationales for which these products work better in terms of cutting down skin and soft tissue infections is because there is a persistent effect or something, and whether the cumulative effect is just that persistence effect magnified, it doesn't make a lot of sense necessarily that you need both of those measures.

I think the problem with these models also is the same problem we face with antibiotic trials. Most clinical trials, like cholesterol, are sort of just the person and the drug. Here we have the wash, the person and the bugs. So, there are three things to look at here. If we are going to look at

CFU reductions, the clearest thing to look at is the preoperative scrub because each person is their own control. Looking at some of the studies were you would look at ten different people and give them five different drugs, the confidence intervals are huge. By taking a mean or median it doesn't make any sense if somebody gains a log when they wash their hands and somebody else loses five logs. I don't think the mean or the median means anything.

So, I think whatever we do decide about these trials, we have to make them a lot tighter and make the analysis a lot more logical. For instance, if soap and water is our control, so our placebo effect, and the others don't go beyond placebo how do we get a delta? I mean, what do we do in that sort of situation? Tom, you may want to talk about that.

Finally, I think there is a difference between resident pathogens and transient bacteria. Those two questions have to be answered separately because looking at the resident bacteria from that

slide that John showed, and also knowing the history of having to be greater than 10 5 CFUs per gram of tissue to create burn infections, it may be different for certain pathogens, but I think there probably is more likely to be a sigmoid curve than a continuous curve.

DR. WOOD: Tom, do you want to respond to that?

DR. FLEMING: Well, Dr. Leggett raises a really key question here among many of his important comments. One of the questions was whatever we use for our control, soap and water, whatever it is, can we use a non-inferiority margin? I think one has been proposed here of saying you have to rule out 20 percent of the effect.

First of all, we have to be very clear about what the effect is in the active comparator. Secondly, we are doing two things at the same time. We are using a surrogate endpoint which is reductions in log, and we are using non-inferiority trials where we are saying how much can we give up

before it is clinically meaningful? I often call the combination of surrogate endpoints and non-inferiority trials my worst nightmare because in most cases I don't have confidence in either one. I don't have confidence that we know the surrogate is reliable, i.e., you have to know how many log reductions do you have to achieve in order to provide the benefit. Well, to do a non-inferiority margin I not only have to know that, I also have to know the function relationship so well that you can tell me how much I can lose on that before it translates into a meaningful increase in infection. Well, as we know, we don't have data on establishing the surrogate in the first place, so how can you tell me how much you can give up on the surrogate effect before it translates into something clinically meaningful on the clinical effect of infection rate?

Now that I have the mike, can I just follow-up on a different issue that relates to some of the comments? The example that I gave of mother-child transmission of HIV and CD4 not even

being in the causal pathway by which the mother transmits HIV I think is relevant to our setting when we look at some of the examples here. When we look at the perioperative skin preparation, when we look at the skin-stripping research that Dr. Powers was talking about, bacterial levels on superficial skin layers may not be the causal pathway; it may be at lower levels.

Dr. Patterson raises the question about the endpoints. She was basically, in my words, saying there may be multi-dimensional components that influence this risk and we may be only dealing with one component. This is reminiscent to me of severe sepsis discussions where we have multiple organ failures and we can go after one of those components and people are complaining about don't ask me to improve survival because I am only dealing with one component. Well, if I am dealing with only one component and that is not sufficiently multi-faceted to translate into clinical benefit, then the truth is I haven't achieved clinical benefit. So, I have to do those

trials to find out whether or not this intended biologic effect translates into truly meaningful clinical benefit.

DR. WOOD: And we do know that antibiotics administered prophylactically had a profound effect here. So, in spite of all the other variables that are in play--different surgeons, different everything--they seem to be pretty dramatic.

DR. FLEMING: Could I ask one question?

DR. WOOD: Yes.

DR. FLEMING: Dr. Boyce, in your second to the last slide you had made the comment that there are no published data demonstrating the cumulative activity of healthcare personnel handwash agents and lower rates of infection. Are we saying here that absence of evidence is evidence of absence? I am wondering is there actual data that would establish that we don't have--what I would really be interested in is not is there absence of evidence but is there evidence to indicate that cumulative activity doesn't result in lower infection rates.

DR. BOYCE: I am not aware that there is any evidence of that nature either. I don't think anyone has looked at the issue of cumulative

activity to determine whether it does or does not impact on infection rates. When you look at what happens in the hospital, when I go to make rounds in the morning I want whatever I clean my hands with to be working at eight o'clock in the morning, the first wash, and I am not really too concerned whether efficacy is greater on the 10th wash, which is what the protocol calls for, or the 20th or 30th or 40th all in one day, which is what really happens in the real world. The risk of the patients developing an infection isn't related to whether you take care of them after your first wash or after your 10th wash. So, frankly, I just fail to not only see any evidence, I fail to see the logic in requiring a cumulative activity in something that is used 20, 30 or 40 times a day during the work shift.

DR. WOOD: But the evidence that any of the other measures are related to reduction in

infection isn't there either.

DR. FLEMING: Let me just probe that. I think I am going to say the same thing but just to probe the logic, if I follow what you are saying, John, the FDA is saying that with the first wash you want 2-log reduction and with the 10th we want 3, following what you are saying, I would like to have 3 both times. But what they are saying is 2 and 3, and what I hear you saying is 2 is enough at the first wash; we don't need the evidence at the 10th. I would justify that conclusion if you showed me data that indicated that products that give 2 at the first and 3 at the 10th don't give added benefit in preventing infection compared to products that give 2 at the first and 2 at the 10th.

The reality is we don't have data on any of this, but given that we don't have the data on any of this it is hard for me to understand how we can advocate weakening the standard as you seem to be advocating for the cumulative wash.

DR. BOYCE: I just don't think that the

rationale is there for requiring a cumulative effect.

DR. WOOD: Let's move on. We are not going to get more data, I don't think. Terry? And this is the last question before lunchtime.

DR. BLASCHKE: I don't know if it is a question or not. I think we have heard a lot of epidemiologic data, and we have read a lot that certainly supports the idea that both handwashing and perhaps antibacterial handwashing is efficacious. What we don't know is if it is efficacious in every situation. I guess I may be anticipating some of the discussion that we will have this afternoon, and I think it goes along with what you were alluding to, Alastair, and that is that we may need to look at some sort of studies, enrichment studies that really allow practical carrying out of such clinical studies to generate the kind of data that I think Dr. Fleming is talking about. One of the things that I think should be happening internally within the FDA, perhaps with its advisors, is to try to figure

out--you know, rather than looking at population as a whole where large samples might be required, really to look at the clinical situations where transmission via healthcare workers is, in fact, at a higher frequency than we might actually be able--I mean, FDA is faced with trying to regulate, regulate in an even-handed way and on a level playing field way.

DR. WOOD: Let's break for lunch and be back at one o'clock. We have greatly overrun this morning because the talks overran a lot. I have asked Shalini to get us a timer for this afternoon, which I will enforce, and I strongly suggest that the FDA and all the other speakers make sure that they get these talks into whatever the agreed time is. In fact, if there are ways to get these talks reduced, as we have used up so much time this morning, I think we should try and do that over the lunch break. So, let's make sure that you don't overrun and, if possible, underrun because the timer will be running. Let's be back here ready to start at one o'clock.

[Whereupon, at 12:10 p.m., the proceedings were recessed for lunch, to resume at 1:00 p.m.]

A F T E R N O O N P R O C E E D I N G S

Open Public Hearing

DR. WOOD: We will now begin the open public hearing but I must first read the following: Both the Food and Drug Administration and the public believe in a transparent process for information gathering and decision-making. To ensure such transparency at the open public hearing session of the advisory committee meeting, the FDA believes that it is important to understand the context of an individual's presentation. For this reason, FDA encourages you, the open public hearing speaker, at the beginning of your written or oral statement, to advise the committee of any financial relationship that you may have with any company or any group that is likely to be impacted by the topic of this meeting. For example, the financial information may include a company's or group's payment of your travel, lodging, or other expenses in connection with your attendance at the meeting. Likewise, FDA encourages you at the beginning of your statement to advise the committee if you do

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

The first speaker is Dr. Felton. You have fifteen minutes and the next speaker has five minutes.

DR. FELTON: Good afternoon.

[Slide]

The title of my talk I guess is difficult but it is proposal for additional intended uses and performance criteria for the TFM: Topical antimicrobials for skin site preparation prior to the placement of percutaneous medical devices intended to remain indwelling. It is a fancy way of saying essentially that if you put a device through the skin, what performance criteria should you have for the topical antimicrobial.

[Slide]

I am Steve Felton. I am staff scientist for BD, a major manufacturer of topical, pharmaceutical and medical device products.

[Slide]

We have gone over this a lot this morning so I won't go through it, except for the "and"

part. Under patient preoperative skin preps there is a subheading, pre-injection, the 1-log reduction in surrogate endpoint. I would like to propose that we have some kind of performance criterion set down in the monograph which would include the information essentially that there is no worse or non-inferiority performance standard for topical antimicrobials with regard to risk for infection for indwelling devices.

[Slide]

I am trying to make this as quick as I can. Here are some of the examples of the devices that may be included in this category. You have short-term peripheral catheters, central venous catheters, peripherally inserted central catheters, surgical pins, intraosseous infusion devices, epidural catheters for chronic pain management and devices for continuous ambulatory peritoneal dialysis. If you got an earlier version of my

slides, it will have abdominal and that was wrong.

[Slide]

These devices have certain things in common. They all go through the skin and they keep the hole from healing after you put them in an leave them in. These devices can remain in place for as little as a few hours for short-term peripheral catheters to literally a year or more.

[Slide]

These devices have significant risk of infection and there is information to predict the risk as it relates to the time and/or placement of the device. Topical antimicrobials applied to the site prior to inserting the device have been demonstrated to reduce the risk of developing an infection and the relative efficacy of the topical antimicrobials is inversely related to the risk of infection. By the way, these citations on this slide at the bottom are the same ones that Michelle Pearson was referring to earlier in the question and answer session.

[Slide]

I am going to use the special case of catheter-related bloodstream infection just to try to develop my argument for why this is important.

This particular group of vascular catheters is used for administration of fluids, monitoring and collection of blood samples. These devices have a significant risk of infection. In the better hospitals in the U.S. it is usually stated around 3-5 percent. In other institutions in the United States you are talking about 10 percent. Now I am dealing specifically with central venous catheters, of course. You go to Europe and you are talking about 25-40 percent infection rates. They use the products a little differently over there.

These infections are not insignificant. There have been estimates of between 296 million and 2.3 billion dollars per year in additional medical costs to treat and otherwise deal with these infections. Mortality is between 5,000 and 20,000 cases per year.

[Slide]

Topical antimicrobials are critical for

placement of these devices. The major cause of these infections is from skin microorganisms, although I will say that there are minor causes such as infusate contamination and also breaks in the line at the hub, etc. However, the topical antimicrobials are not intended to deal with those.

In these studies that are shown here, they have proposed, especially Maki and Widmer, a large chain of evidence that skin microorganisms not only initiate these infections by colonizing the skin at the insertion site, and these bugs are present there either due to insufficient antisepsis or the bugs are there because the site is recolonized from skin bacteria adjacent.

[Slide]

These microorganisms colonize the subcutaneous and intravascular portions of the device which, if no intervention occurs, can result in a local infection. This can then progress to clinical signs, although in central venous catheters the clinical signs of local infection are not so predominant. Sometimes you can have an

infection that goes straight to catheter-related bloodstream infection with full-blown symptoms at the systemic level, and this does have significant morbidity and mortality and the added healthcare expenses. You are talking about $5,000 to $40,000 in the ICU for each one of these incidents.

[Slide]

The CRBSIs have been studied extensively. I am just going to mention that there are some methods out there that have been developed and independently verified that seem to ways to diagnose catheter-related bloodstream infections, and investigators have shown that the efficacy of the topical antimicrobial can be evaluated in a clinical setting, and the investigators have compared, for example, alcohol with povidone-iodine to chlorhexidine in a number of these studies. These, again, are the same references that Dr. Pearson referred to.

[Slide]

So, in this presentation I have primarily discussed central venous catheters as these devices

are the most extensively studied. These devices have significant infection rates, 3-5 percent at the better institutions, and significant mortality, 5-20 percent of the subjects with clinical CRBSI. These infections are estimated to increase the U.S. healthcare cost by 2.3 billion dollars a year, up to that amount.

However, percutaneous medical devices are all similar in that they remain in the hole through the skin barrier. Therefore, any intended use labeling or performance criteria developed for CRBSI should be applicable to other percutaneous medical devices. Unlike the current performance criteria in the TFM, the efficacy of topical antimicrobials intended to reduce indwelling percutaneous medical device infections can be demonstrated in clinical trials in the intended use population. Therefore, the TFM should identify the need for and establish performance criteria for the clinical evaluation of indwelling percutaneous medical devices. Thank you.

DR. WOOD: Just help me understand, you

are not suggesting--or are you suggesting that you should not do clinical trials?

DR. FELTON: I am suggesting that for this particular indication clinical trials are indicated.

DR. WOOD: Okay. The next speaker is Dr. Ijaz, who is from Microbiotest. He has five minutes.

DR. IJAZ: Good afternoon. First of all, I would like to thank the organizers for providing me this opportunity to express my views on this topic, which is hand hygiene and viral surrogates to demonstrate efficacy of topical agents against viruses.

What I want to raise here is that we have been discussing microbiological surrogates but we have not touched viruses and that is what I want to raise. I have only five minutes so I will just make my point very briefly. We know the significance of viruses, and viruses in general continue to emerge and re-emerge. If one looks at the past 30 years, we have seen from the '70's a

focus on enteric viruses, hemorrhagic fever viruses, and in the 1980's retroviruses and in the '90's, you know, sin nombre and more hepatitis viruses, and more recently we have seen influenza virus and SARS emerge.

So, the importance of viruses, from a morbidity and mortality point of view, globally is well documented, and these viruses continue to emerge. Specific to this meeting, in the U.S., 5 percent of nosocomial cases are due to viruses and greater than 32 percent are in the pediatric wards, of which RSV is the most common.

Hands play an important role in spread of many virus infections and proper handwashing by care givers and food handlers for interruption of spread of viruses and other type of pathogens is universally recognized. This has been demonstrated in intervention experimental studies, as well as studies conducted in the clinical setting, particularly dealing with the rotavirus infections and rhinovirus infections. Infectious viruses have been recovered from naturally contaminated hands.

As a case in point, I can document here these studies dealing with hepatitis C virus, RSV, rhino and rotaviruses.

Now, although the FDA's Center for Food Safety and Nutrition recognizes the significance of viruses being disseminated by food handlers and healthcare workers, the role played by hands in this regard in the TFM has not been addressed, and that is the issue that I want to raise.

Proper antiseptic procedures for use for decontamination of hands can interrupt such disseminations. The question is do viruses survive on hands? We looked, in a very simple, small experiment, at the survival of rhinoviruses and BVDV which is used as a surrogate for hepatitis C on finger pads contaminated with these viruses. Of course, all of these studies that I am reporting here, they have gone through IRB approval. You can see that both of these viruses may survive well on the finger pads of human subjects for 20 minutes. Studies done at the University of Ottawa indicate that some naked and some animal viruses survive

more than an hour on hands.

Here is a commercial from CDC, which we saw in the morning session as well. When we are thinking about testing topicals and their activity against viruses, there are a number of methods which are out there, and I am picking the one which I believe is better than the other ones to demonstrate efficacy of these products. The methods that I am referring to have been peer reviewed. The data generated by these methods have been published in peer reviewed journals and these methods are also the ones that have been approved by ASTM.

I am not going to go into details of this method which deals with the use of finger pads to study the efficacy of the products.

DR. WOOD: I am afraid your time is up. Let's move on to the next speaker, who is Dr. Osborne, from the FDA.

The Quest for Clinical Benefit

DR. OSBORNE: Good afternoon.

[Slide]

I am Steve Osborne, a medical officer in the Division of Over-the-Counter Drug Products. I have shortened my presentation per request of the

Chair. You will find all the slides in the handout. I have also shortened how much I am going to speak about each slide. If there are any questions, I will be available later. The title of my presentation is the quest for clinical benefit.

[Slide]

We have heard Tia Frazier and some other members mention that obtaining clinical data from clinical trials of healthcare antiseptics can be a daunting task. Two of the issues that we face at FDA in evaluating healthcare antiseptics for the monograph are do clinical trials assessing infection rates provide definitive evidence of clinical benefit?

[Slide]

And, does the clinical evidence link surrogate endpoints with clinical benefit?

[Slide]

First I would like to run through the

major categories of healthcare antiseptics and give a quick example of each. The alcohol symbolizes ETOH or IPA for isopropyl alcohol found in a Purell handrub or Purell instant sanitizing handwipe. Chlorhexidine gluconate, or CHG, is found as 2 percent or 4 percent. The trade name is Hibiclens or Hibiprep--iodine or iodophor--we all know PI or betadine. Triclosan is found in the Gojo antimicrobial lotion soap.

[Slide]

The quaternary ammonium compounds, as an example there is benzalkonium chloride, known as Zephiran. Chlroxylenol is found in the wash and dry towelette; and triclocarban is found in the common Safeguard soap.

[Slide]

I won't dwell on this slide but it shows the antimicrobial spectrum of the common antiseptic categories. It is from the CDC 2002. What the slide shows is that the antimicrobial spectrum is broad for most of these products, except for gram-negative activity with the phenols and

gram-positive activity with the quaternary ammonium. The time frame fast, intermediate or slow is not exactly defined but for fast you can think of as seconds; intermediate as seconds to minutes; and slow as minutes to hours.

[Slide]

The citizen's petition and comments were submitted to FDA in 2001 and 2003 by the industry coalition made up of the Soap and Detergent Association, or SDA, and the Cosmetic, Toiletry, and Fragrance Association, or CTFA. A citizen's petition is the process whereby the public or someone can ask that FDA change the monograph. The coalition submitted references and requested that FDA lower the efficacy standards.

[Slide]

Two broad categories were encompassed by 155 abstracts and articles. They were invasive procedures such as surgery, or non-invasive procedures such as using a handwash to reduce nosocomial infections.

[Slide]

Of the 155 articles and abstracts, 58 percent covered handwashes; 26 percent were patient preop preps; and 16 percent were surgical scrubs.

Overall, the weight of the evidence of clinical benefit was not persuasive for changing the current efficacy criteria. As a common thread, there was no link between surrogate endpoints and infection rates.

[Slide]

This is a summary of some of the limitations in these studies when you look at them in the context of our monograph process. Not each study had each limitation; some had more than one. The common thread, as mentioned, was that surrogate endpoints were not correlated with the clinical outcome. Some of the studies were not randomized. They might have gone back 30 or 40 years in some instances. A placebo was not used in some of them or a control. On occasion they were retrospective, without a comparator or whatever happened before that period of time.

Multiple confounders might have been

present. You can think of that as when you introduce a new healthcare antiseptic, for example a handwash, but at the same time introduce a training program involving posters, reminders, brochures, etc., such that when you later try to look at infection rates you are not sure if what you have done is simply helped the infection rate with the antiseptic or whether you have changed the behavior of the subjects in the test.

Inadequately powered, and we will see that in one of the studies by Luby. No statistics. That is not so common in the last few years but it does make it difficult to test your hypothesis if that is the case. Lack of standardization of product use--this is complicated. When you introduce, for example a handwash, you don't always have the capacity to regulate how much of the handwash people use, nor how long they use it in terms of the washing cycle. Irregular patterns of data collection--one study looked at 26 hospitals using a healthcare antiseptic and only 13 returned data for later analysis.

Failure to address the TFM indication. This is a complicated thing but if the study is looking at something that is not specifically the

way the TFM has the indication for the handwash patient preop prep or surgical scrub, then we cannot use that study in making a regulatory decision. Examples would be if a healthcare antiseptic is used in acne applications or, as we heard earlier, in a patient preoperative shower, which is not a TFM indication.

I am going to show some examples of studies from the industry coalition and from three literature reviews performed at FDA. I would like to emphasize that these studies are notable examples trying to analyze the answers to important clinical questions. They are not being criticized. However, for one set of the study or other they have a limitation where, by the design of the study, we at FDA are not able to use the results from it to make a regulatory decision for the monograph.

[Slide]

Maki et al., in 1991, looked at catheter infections and Luby et al., in 2002, looked at impetigo. First the Maki study.

[Slide]

It was randomized, unblinded study in 668 subjects with IV catheters, all of which were

central venous or atrial. Two percent chlorhexidine gluconate was compared with 10 percent povidone-iodine and 70 percent isopropyl alcohol. The agents were applied before insertion of the catheter and then every 48 hours thereafter until the catheter was removed.

[Slide]

When the catheter was removed, endpoints looked at were the local infection rate and bacteremia. For the local infection rate, it was designed as greater than 15 colony-forming units at the catheter tip upon removal, and that is synonymous with catheter colonization. The infection rate locally was 2.3 percent for CHG versus 7.1 percent for alcohol and 9.1 percent for PI, and that was statistically significant in favor

of chlorhexidine gluconate.

The harder endpoint of bacteremia had a total of 10 cases out of the 668 catheters. This is a rare occurrence, in other words. One was found with CHG, 3 with alcohol, 6 with povidone-iodine, and the difference was not significant. As you can see, when you have a low incidence of an endpoint it is difficult sometimes to show a difference between products.

[Slide]

However, there was no correlation between reduction of bacteria at the site of the catheter insertion with the resulting infection rate in the individuals receiving the catheter, and therein lies the limitation if you try to apply it to what we need at FDA.

The application of the antimicrobial post-catheter insertion limits the ability to relate to a monograph application, which is to apply the product, insert the catheter and then perhaps later simply to look at infection rates. Applying it every 48 hours confounds the result.

[Slide]

Luby, I am going to pass over because of time.

[Slide]

Dr. Michelle Jackson, who we heard from earlier, performed a literature review on surgical hand scrubs. Over 300 articles were screened for clinical benefit. None conclusively linked reduction in bacteria with reduction in infection rates.

[Slide]

Examples are Bryce et al., 2001, that looked at a 70 percent isopropyl alcohol leave-on product in 70 scrubs by surgeons, the people who know how to do the scrubbing. This was an in-use hospital evaluation and 14 mL of the product was used over 3 minutes and compared to 4 percent CHG and 7.5 percent PI in reducing bacteria. The endpoint was postop bacterial counts on the hands of the surgeons. No infection rates were studied in the patients.

[Slide]

Parienti et al., 2002, performed a hand-rubbing with alcohol leave-on solution and looked at the 30-day surgical site infection rate later. This was a randomized, crossover equivalence trial comparing the 75 percent alcohol leave-on product with the standard 4 percent PI and

4 percent CHG as surgical scrubs. Six surgical services and 4,287 patients were looked at.

[Slide]

The surgical site infection rate was 2.44 percent with alcohol versus 2.48 percent with the combination of PI pus CHG. That was not significantly different. The scrub time compliance was better with the alcohol rub. So, that goes along with what some other people have said, that the alcohol might be better tolerated. Surgical site infection microscopic was not provided, and the surgeon who reported the surgical site infection in the patient was not blinded.

[Slide]

Another member of our division, Dr. Collen Kane Rogers, performed a literature review of

healthcare personnel handwashes from 1994 to 2004, and 222 studies were reviewed for clinical benefit or efficacy. None showed a definitive link between bacterial reduction and reduction in infection rates.

[Slide]

An example of an interesting study is Swoboda et al., 2004. This was a 3-phase, 15-month evaluation incorporating an electronic monitor, that is, to see if the patients were actually washing their hands and then to voice prompt and remind them to do so. So, approximately a 6-month monitoring period, followed by voice prompt, and then a monitoring period was conducted. Compliance with handwashing improved by 35 percent in the second phase versus the first, and by 41 percent in the third phase versus the first. Patients were colonized--not necessarily sick but colonized by either methicillin-resistant Staph. aureus or vancomycin-resistant enterococcus in 19 percent of the initial phase, 9 percent of the second, 11 percent of the third phase, indicating that perhaps

there was a trend towards lower colonization. Again, you don't know though whether this is a change in behavior but that is what the study was looking at.

[Slide]

Another member, Dr. Peter Kim, from the Division of Anti-Infective Drug Products, looked at 400 articles in the patient preoperative literature, and in this review searched for bacterial log reduction data post scrub compared with pre-scrub, and then in the same article, searched for surgical site infection rates.

[Slide]

The majority of these studies were performed in animals and that answers the question brought up by a panel member earlier. None of these studies found a link between colony-forming units of bacteria in surgical site infections.

[Slide]

A secondary topic looked at in this review addressed the question is there a minimum number of bacteria in a wound that predisposes to infection?

This is a 100,000 bacteria or 10

5 rule that we have

all heard about through the years. Of course, this

may vary with the type of bacteria, 10

5 Staph. epi.

is not the same, of course, as even 100 shigella.

[Slide]

On this threshold for infection, Kass, in '57, looked at 2,000 patients for pyelonephritis and found all of them had over 100,000 bacteria, and a similar thing with UTI patients.

Krizek, in 1967, showed a 94 percent graft success rate when the pre-graft bacterial count was less than 100,000/gram of tissue. That was brought up earlier by a panel member. And, the rate would go as low as a 20 percent graft success if there were more than 100,000 bacteria/ gram.

[Slide]

From this review, Cronquist et al. has an interesting study in 2001 of 609 neurosurgery patients undergoing craniotomy or ventriculo-peritoneal VP shunt. This study looked at pre-scrub and post-scrub bacterial counts from the head and the back.

[Slide]

From the head, pre-scrub was 4.13 log and from the back 2.39 log of bacteria. Post-scrub was

0.63 and 0.54. The agents used in this study were PI scrub followed by isopropyl alcohol wipe off and then a PI paint.

[Slide]

Twenty surgical site infections were noted, 19 from the craniotomies, and these involved mostly staph. species and Propionibaceterim acnes. No correlation was found between the pre-scrub or the post-scrub counts in surgical site infection rates. Remember from that slide that all of these counts were less than 105.

[Slide]

So, I return to two key issues FDA faces, do clinical trials assessing infection rates provide definitive evidence of clinical benefit?

[Slide]

And, does the clinical evidence link surrogate endpoints with clinical benefit? These are issues for the panel to discuss.

I would like to next introduce Dr. Thamban Valappil, from the Office of Biostatistics in the Division of Biometrics III, who will discuss some of the statistical issues.

DR. WOOD: Thanks very much for getting that done so quickly.

OTC-TFM Monograph Statistical Issues

of Study Design and Analysis

DR. VALAPPIL: Thank you, Dr. Osborne. Good afternoon.

[Slide]

I am Thamban Valappil, statistician in the Division of Biometrics III.

[Slide]

Now I will go over some of the statistical issues and limitations of the study design and analysis in the OTC TFM monograph. The outline of my presentation is as follows: introduction; summary of statistical issues; current TFM trial design and analyses with surrogate endpoints; statistical issues of study design and analyses; options for trial design and efficacy criteria

using surrogate endpoints.

[Slide]

Introduction--previous presentations on issues involved in validating surrogate endpoints, in the absence of clinical trials data, FDA still needs to address current products under review. This talk discusses issues related to analysis of data obtained on surrogate endpoints. It does not address clinical relevance of statistical findings or differences in analysis of data based on surrogate endpoints.

[Slide]

Now I am going to discuss briefly the summary of statistical issues. The primary endpoint is the log reduction in bacterial counts from baseline. It is a surrogate endpoint and its clinical relevance has not been validated, as I said earlier.

Data analysis and variability issues--there are a couple of different ways we can look at the data. One way is using the binary endpoint, which is the percent of subjects who meet

the threshold log reduction and the other one is using log reduction in bacterial counts. However, in each of them there are advantages and disadvantages.

Log reductions are continuous, numerical data with relatively large variability. The current TFM recommends mean as the measure to analyze the spread in the data. However, median would be another possible option although it is not mentioned in the current TFM.

Study design and controls--currently, a non-comparative study design has been used in which the test product is not directly compared to the active control. Vehicle and active controls are mentioned in the current TFM, however, the role of these controls is not well defined.

[Slide]

This table shows a brief layout of what is available in the current TFM. Use of various controls is mentioned under the surgical hand scrub section of the monograph. But for preoperative skin preparations and healthcare personnel handwash

only active control is recommended.

For comparing the mean log reductions t-tests are recommended. Under preoperative skin preparations, a confidence interval approach based on the difference in success rates between the test product and the active control has also been documented.

However, it is important to note that in the current TFM the efficacy criteria do not use any of these statistical tests, except using the mean log reductions to meet a threshold value. The last column displays the sample size required for each of these documents.

[Slide]

A brief layout of the current TFM recommendations are as follows. TFM currently recommends randomized and blinded trials, also recommending use of active, vehicle and/or placebo controls. However, in the current TFM a non-comparative study design is used in which the test product is not directly compared to the active or vehicle control. Mean log reduction meeting the

threshold log reduction has been used to demonstrate efficacy.

[Slide]

Although vehicle and placebo controls are mentioned in the current TFM, the majority of the NDAs only have test product and active control arms. Active controls have only been used for internal evaluation of the study methods. Efficacy assessment does not include a direct comparison of test product performance to active control, vehicle or placebo controls.

[Slide]

Statistical issues of study design and analysis--currently, the TFM recommends using log reduction from baseline as the primary endpoint and it can be influenced by few extreme observations. As a suggestion, we could discuss median log reduction as another possible option. Median is less sensitive to extreme log reductions or outliers. It is shown here in parentheses as the current TFM does not specify it.

[Slide]

The efficacy criteria in the current TFM are based on point estimates and do not include confidence intervals to evaluate variability.

Consequently, a few extreme observations can potentially drive the efficacy results.

[Slide]

Now let us look at this figure which shows the log reduction in bacterial counts using the threshold approach. This is just an example to illustrate the potential problems if the variability of the data has not been considered. Here the threshold is set to logs, as marked by the blue dotted line. There are 18 subjects and 14/18, 78 percent, of the subjects, marked in red, have failed to meet the threshold. As you can see, only 4 subjects, marked in blue, are basically driving the results to meet the required log reduction. Instead of mean, if we use the median, which is 1.7 log, this study would have failed to meet the threshold log reduction.

[Slide]

Now let us look at a few examples to

illustrate the importance of controlling variability and the roles of active and vehicle controls. In this figure, for illustrative purposes, if we look at the point estimates, as done based on the current TFM, the test product may seem better than the active control however, when we consider variability, the confidence interval for the test product and the active control overlaps, as you can see in the next figure.

[Slide]

As you can see here, the confidence intervals for the active control and the test product overlap and both are better than vehicle. As Dr. Powers has pointed out, it is not how you define the threshold but how you analyze the that data is important.

For simplicity, in this figure the confidence intervals of the individual products are displayed rather than the confidence intervals around the treatment differences. It should be noted that demonstrating superiority in this situation is a mechanism to control variability but

that does not address the issue of clinical relevance. Let us take another example.

[Slide]

Here the confidence interval for the test product and the active control overlaps and it meets the threshold based on the current TFM. However, if we introduce the vehicle control the test product appears no better than vehicle. Therefore, it is important to incorporate a vehicle or placebo control in the trial design.

[Slide]

The current TFM has recommended using binary outcomes, however, the efficacy criteria are not based on binary outcomes. Accordingly, a subject will be classified as a success or a failure based on meeting the threshold log reduction.

These are advantages and disadvantages in using this approach. The advantages are that the outcome will be centered on number of subjects and not on organisms, which provides greater confidence that it is meeting the threshold. Also, the effect

of variability will be reduced. However, one disadvantage will be that this method does not differentiate the magnitude of log reductions among those who meet the criteria for success.

[Slide]

Let us look at this example. In this figure, based on binary outcome, 90 percent of the subjects, marked in blue, meet the threshold reduction and provide greater confidence that it is meeting the threshold compared to the small chart, as you can see in the upper left-hand corner, in which only a few subjects meet the threshold.

[Slide]

Now let us consider one of the agency approved study data. This table is based on an NDA approved for surgical hand scrubs. All met the required log reduction except for active product number 2 on day 5. Also, the success rates widely vary among the 3 products and mask the difference among the median and mean. On day 2, if you notice, the success rate goes from 100 percent for the test product to 45 percent for the active

control product number 2, as highlighted. however, they all meet the required mean log reduction. You will also notice that if the success rate is higher, mean and median does not make much of a difference. But if the success rate is low, the median is much more conservative since it is not influenced by extreme outliers.

[Slide]

Sample size issues--in the current TFM sample size is estimated based on allowing a test product to be as much as 20 percent worse than the active control in the mean log reduction. However, the basis for the 20 percent margin is not clearly stated. Majority of the current submissions do not follow the recommended sample size as specified in the TFM and only use a sample size of about 30 subjects per treatment arm.

[Slide]

There are several issues that need to be addressed before the design and efficacy criteria are discussed. The various issues are, issue number one, how to analyze the data obtained on the

surrogate endpoint of log reductions in bacteria.

Issue number two, how to take into account the variability in the data collected when measuring effect of the product.

Issue number three, how to take into account the variability in the test methodology.

[Slide]

Now let us go through the issues in detail. The first issue is how to analyze the data obtained on the surrogate endpoint of log reductions in bacteria. There are three options, mean, median and percent of subjects who meet the threshold. Please note that these are all for discussion.

As we know, mean log reduction can be easily influenced by extreme observations. However, median log reduction is less sensitive to outliers or extreme observations. For percent of subjects who meet the log reduction criteria the outcome is centered on number of subjects who meet the threshold and may provide incentive to study conditions of use that provide highest success

rates. Also, it provides greater confidence that it is meeting the threshold.

[Slide]

The next issue is how to take into account the variability in the data collected. There are two options. Option one, we can examine the outcomes as defined on the previous slide with a threshold for lower bound of the confidence interval. There is a pro and con in using this method. The pro in using this will be an improvement over examination of point estimates alone. The con is that it does not take into account the variability in the method.

The second option is to examine confidence intervals around the treatment difference between the test product and some control. Her the pro is that it allows for examination of variability in the methodology across treatment arms. The con is that it may require a larger sample size for products with lower success rates.

[Slide]

Issue number three, how to take into

account variability in the test methodology. There are two options. Option one is equivalence or non-inferiority showing that the test product is no worse than the active control by some clinically meaningful margin. The pro is that it allows for comparison with an active control treatment to rule out loss of effect relative to active control. The con is that it lacks constancy of effect of active control in previous studies, possible overlap of effect of active and test product with the vehicle and, hence, no basis to select a clinically meaningful non-inferiority margin.

The other option is to test for superiority of test product to the vehicle and superiority of active control to the vehicle. The pro is that given lack of constancy of effect with both active control and vehicle control, it allows internal validity of comparisons. The con is that it may require a larger sample size than current TFM standards. How large a sample size will depend on the product efficacy over the vehicle.

[Slide]

Controlling variability in test methodology--to address these issues, let us consider a 3-arm trial design which includes the

vehicle, the active control and the test product. It is important to note that the test product and active control both demonstrate superiority to the vehicle. Also, it is important to note that there are multiple sampling times and, accordingly, there is multiple hypothesis testing involved. The superiority of the test product will be demonstrated only if all tests are statistically significant.

[Slide]

This figure shows the sample size requirement for the superiority test over the vehicle using a binary endpoint. As success rates increase, as you can see in the figure, and the treatment difference over the vehicle is large, the required sample size is much less.

For example, if the success rate for the test product is 90 percent and the treatment difference compared to vehicle is 10 percent, then

a sample size of 199 subjects per treatment arm is required. Similarly, for a 20 percent treatment difference, 62 subjects, and for a 30 percent treatment difference, 32 subjects are required per treatment arm. Therefore, the message is that more effective products require smaller number of subjects.

[Slide]

With this, I conclude my presentation and thank you for your attention. Now I would like to thank Dr. Daphne Lin, Statistical Team Leader and Acting Deputy Director of the Division Biometrics III, for her valuable contributions. Thank you.

DR. WOOD: Could you put slide 13 back up? I don't understand why you would ever want to do a non-inferiority trial for a surrogate like this. I mean, surely you would always do it against vehicle.

DR. VALAPPIL: I am not proposing a non-inferiority trial. This is just an example to illustrate--

DR. WOOD: Yes, I mean, the reason you

normally would do a non-inferiority trial is where it would be unethical to do a study.

DR. D'AGOSTINO: This is not a non-inferiority. The active is just for internal validation. The active doesn't have to be compared against the test.

DR. WOOD: Oh, I see.

DR. D'AGOSTINO: It is confused I think the way he has it, but isn't it just--

DR. WOOD: Let me rephrase the question. It seems to me there is no justification for ever not doing a study in a surrogate where you don't have just the vehicle as the control. All these numbers on your last slide look pretty trivial to me given the numbers we see in other studies, and this is a very easy study to do so I don't see what the issue is here.

DR. FINCHAM: Alastair, may I ask a question?

DR. WOOD: Yes.

DR. FINCHAM: On your slide 16 you go through study number 1. Is this hypothesis data?

DR. VALAPPIL: No, this is real data. This is the data collected from one of the NDAs we have approved.

DR. FINCHAM: Is it confidential or is it not referenced because of that?

DR. VALAPPIL: I cannot address the study.

DR. WOOD: So, where is the vehicle control there? DR. VALAPPIL: Actually, number two is the vehicle control; but it is not actually vehicle.

DR. WOOD: That is a study you received that didn't have a vehicle control in it? Is that right?

DR. VALAPPIL: The purpose of this slide is to show you the difference in the mean and median, and also to find out the difference in the success rates.

DR. POWERS: You are pointing out an important point, there are no vehicles in these and that is what Dr. Wood is actually asking.

DR. WOOD: I thought that was the question I was asking and I am getting a very confused

answer. Are you looking at studies here that do not contain vehicle control? Yes or no? Yes. Is that right?

DR. D'AGOSTINO: But can I ask a question? Are you suggesting that in the future studies should be done with the real vehicle, or are you saying that what you are calling a vehicle is somehow or other a low-level active?

DR. VALAPPIL: No, no, that is not what we are proposing, but I think it would be better to have the vehicle incorporated in the trial design so we know what is the product effect compared to the test product.

DR. WOOD: Put up slide 7 again. As I read what you have there, it says the current TFM--maybe I am reading it wrong--recommends that you can do a study just with active control. Am I reading that wrongly?

DR. VALAPPIL: No. What I was trying to tell you is that--

DR. WOOD: No, wait, are we reading that wrongly? Can you do a study right now with just

active control?

DR. JOHNSON: Yes.

DR. WOOD: Yes is the answer.

DR. D'AGOSTINO: But you don't have to contrast the active with the test. You ask the question does the active exceed the threshold and, if it does, you say you have internal validation. Then you ask does the test exceed the threshold, and you never make the comparison of active with the test. Is that right?

DR. JOHNSON: That is correct also.

DR. WOOD: I guess that is the point I am making, it is crazy.

DR. D'AGOSTINO: Can I just jump in here? If you do a test where you have the vehicle, the active and the test, you look at the active versus vehicle; you look at the test versus vehicle; and you hope both of those are significant. At that point, you still also need the log reduction for the clinical, but we don't know what clinical significance means because we don't know how to tie it, but that would be one possibility. Then you

would have to do that for every single time period.

DR. VALAPPIL: Right.

DR. WOOD: We can take questions for all of these now. Any other questions?

DR. FINCHAM: I don't think our speaker ever got the chance to answer the question that was asked. Could he do that?

DR. VALAPPIL: Yes, what was the question, please? DR. FINCHAM: Well, I think everybody else answered the question that was meant for you but I don't think you answered the question. I don't think he had a chance.

DR. VALAPPIL: If you can repeat the question I will be able to answer that.

DR. WOOD: Which question? Sorry?

DR. FINCHAM: Well, I think that you both have dealt with it and you referred to the slide that is up there now, and I just didn't know whether you agreed with what was answered.

DR. POWERS: Can I help with this? There are several options within the TFM as to what you can do. Believe me, it is confusing to us too. In

the statistical section of the TFM it states that you can do essentially what is a non-inferiority trial based on a surrogate endpoint with a 20 percent margin. In other places in the TFM it states that you just need to meet a log reduction.

So, what it really does is present you with several options. There is also one part in the TFM that says you can also use vehicle but it doesn't tell you what to do with the information and the vehicle. So, if it is confusing to you, it is because it is confusing and there are several options put out there and it does not specify which one you should use.

DR. WOOD: It is always reassuring to not be uniquely confused I guess. All right, any other questions?

DR. SNODGRASS: I just have a brief comment. It sounds like we should go back to the "paperwork reduction act." You know, you just go back to the drawing board and get rid of the past TFMs and you start over.

DR. WOOD: It is two o'clock; don't get

too ambitious!

[Laughter]

All right, let's move on to the next speaker, and the next two speakers are going to present the industry's view, and the first speaker is George Fischler, and we are generously going to give each of you 23 minutes, which is one minute more.

Industry Presentation

The Value of Surrogate Endpoint Testing for Topical Antimicrobial Products

DR. FISCHLER: And just to start this off, how do you think we feel?

[Slide]

Good afternoon. I am George Fischler, the manager of microbiology for the Dial Corporation.

[Slide]

Today I am speaking on behalf of the Soap and Detergent, and Cosmetic, Toiletry and Fragrance Association Industry Coalition. The SDA/CTFA coalition has previously submitted several detailed comments and has had extensive interchange with FDA

in response tot he June 17, 1994 tentative final monograph, the TFM, for healthcare antiseptic drug products. I will be speaking on the value of surrogate endpoint testing. I will then be followed by Jim Bowman of Hill Top Research, who will talk on statistical issues. We will then be happy to answer any questions.

[Slide]

During this time, the science surrounding topical antimicrobial skin antiseptics has continued to advance. Much of the original analysis done on the use of healthcare antiseptic drug products was developed in the 1970's. Both infection control practice and test methodologies have undergone changes, and the testing and evaluation of these products must be done in the light of current practice.

The coalition has been at the forefront of much of this evolution. While the basic perspective of the coalition has not fundamentally changed since 1995, we believe that our current position and recommendations, updated to include

new information, data and further validation of test methods outlined in the TFM, are well-grounded in the latest science. Our recommendations do not represent a lowering of efficacy standards but, rather, matching surrogate endpoints with current practice, and this is a very important point. We appreciate the opportunity to summarize our perspective and look forward to continuing dialog towards finalizing a monograph that establishes appropriate test methodology and performance criteria representative of a threshold of clinical effectiveness for this important category of healthcare drugs. Our presentation will cover the following topics.

[Slide]

A basic premise of the monograph system is that certain, well-defined categories of drug products that have been determined as safe and effective may be marketed without FDA pre-approval, as compared to the NDA system which requires that individual formulated drugs undergo separate review and approval prior to marketing. A key challenge

of the monograph that addresses healthcare antiseptics is the determination and demonstration of efficacy for a category of drug products that encompasses several distinct active ingredients across a range of indications.

[Slide]

Our first key point is that definitive randomized and controlled clinical trials, typically used to assess therapeutic benefit are not practical in measuring the prophylactic benefits of topical antimicrobial products.

[Slide]

Investigators in this area have stated that definitive, classical, prospective, randomized and controlled clinical trials typically used to assess therapeutic benefits are not practical in measuring prophylactic benefits of antimicrobial products.

[Slide]

Human clinical trials have a number of issues that can blur any potential efficacy result and can cause the size of the study to become so

large that it is impractical, impossible or unethical to conduct. For example, the incidence of infection should be directly related to a specific dose of organisms that causes a particular infection. We have heard a lot about that today. Nmerous mitigating factors influence whether an infection can become established, including the immunological status of the host, the route of infection, direct or indirect transfer of the infectious agent, etc., and we heard a lot more of these confounding factors here today.

In addition--and this, again, is a key differentiator particularly of handwashing--the primary target of antiseptic handwashing is not the individual using the product. Rather, it is to prevent the transmission of pathogens within a relatively large specific population, healthcare providers, thus improving public health. Within that context, many factors not directly related to the efficacy of the product must be considered, primary amongst them being compliance. It is paramount in the development of antiseptic

handwashes or rubs that acceptance, whether through convenience or mildness, is always an important consideration when formulating such products. Manufacturers have made significant improvements in dispensing systems, product forms such as foams, and the mildness profile of products meant to be used repeatedly. In addition, many manufacturers have sponsored studies aimed at looking at ways to improve hand hygiene compliance.

All of these factors make it difficult--and I think that is an understatement--to calculate the level of bacterial reduction needed to demonstrate the benefit from the use of primarily prophylactic agents. For these and other reasons, alternatives to classical, prospective, randomized and controlled clinical trials must be used for evaluating these topical antimicrobials.

Fortunately, there is a substantial body of scientific evidence that demonstrates the public health and clinical benefit of using topical antimicrobial products in healthcare settings.

Such a benefit has been demonstrated repeatedly through studies of bacterial transmission and infection rate reduction. These data allow for determination of effectiveness by benchmarking current antimicrobial products.

[Slide]

Our second key point is that standardized, defined and peer-reviewed test methodologies ensure reliability, reproducibility and comparability of test results. For the purposes of a monograph, it is necessary to establish efficacy methodology and criteria that ensure effectiveness of topical antiseptics. Surrogate testing provides such a methodology. Such testing encompasses both in vitro and in vivo methodologies, and extensive comments have previously been submitted to the FDA on their validity. We shall be presenting some of these data from the published literature, and some of these will be repeats of what you heard so I will jump through them rather fast but there are some key points to bring out from them. It is apparent that over the years many different and

incomparable test methods have been used to assess effectiveness. The efficacy of topical antimicrobial products can be defined as the prevention or reduction of risk of bacterial transmission.

[Slide]

The FDA, in 1978, found that the reduction of the normal flora, both transient and resident, has been sufficiently supported to be considered a benefit. The only determination that remains, therefore, is how much of a reduction in microbial flora will be required to permit claims for the various product classes.

Thus, the agency has previously embraced reduction of skin flora by a prespecified amount as a valid surrogate endpoint for the efficacy of topical antimicrobial products in a clinical setting. Healthcare personnel handwashes or waterless hand rub preparations are largely designed for the removal of transient microorganisms from the skin. These products are used in a clinical setting in an uncontrolled

manner, with little regard for the dosage, the amount applied during handwashing, exposure time, repeat interval, or the amount of water used if the product is intended to be used with water.

Due to the nature of product use, demonstration of efficacy in these products in an actual use setting would be, by definition, uncontrolled and, therefore, poorly suited for study by classical methods. Therefore, these products are tested in a controlled manner by procedures such as the ASTM Healthcare Antiseptic Handwash test, the E1174, or in Europe by the EN-1499 and EN-1500 handwash and hand rub methods that similarly employ surrogate endpoints.

[Slide]

Although the basic ASTM E1174 framework has been in use for many years and has served as the basis for approval of many currently marketed NDA products, researchers have modified it, and we have heard a lot about that, and the method itself has undergone rigorous review within ASTM and several improvements to minimize test variability

have been instituted. The importance of complete and immediate neutralization of active ingredient is foremost among these changes. Incomplete or delayed neutralization can have the effect of overestimating ingredient efficacy. This is shown by a study that looked at a direct comparison of test versions.

[Slide]

The test versions were the current ASTM method, as it is published in ASTM, the ASTM method as it was published prior to 1994, which is a method that was used for many of these NDAs, and the method as published in the 1994 TFM. I will compare three primary parameters, inoculum application, neutralization and timing of the baseline enumeration.

I am going to take a little time to go through this slide because I think this is very important to understand. In the first column we have the inoculum addition. The current ASTM method calls for applying the inoculum to the hand in 3 1.5 mL aliquots. This is the culture of

Serratia marcescens. That is done in order to minimize variability in the baseline because it is very difficult to keep 4.5 mL or 5 mL of liquid in the hand without spilling some into the sink. So, applying it in smaller amounts helps give you a baseline that is much less variable.

The timing of the baseline measurement--this is particularly important when it comes to the 1994 TFM method as written. As you heard Michelle Jackson talk this morning, the way the test is done is that a cleansing wash is performed to familiarize the subjects with the wash procedure. Following that, the hands are inoculated with the Serratia. In the 1994 TFM, as it is written, it is then followed by another cleansing wash and after that the baseline is then calculated.

The way the ASTM method reads is that the baseline is taken following the familiarity wash and then the inoculum. You can see the result that that has in reducing the baseline by almost 3 logs. So, you are starting at a very different point with

the TFM method than you are with either of the ASTM methods.

Again, neutralization--a very important point because, again, the goal of this test, of any test, is as good as it can be to mimic what goes on in real life. I think we would all agree that ultimately the answer is that no test can mimic what goes on in real life but you have to try and minimize the variability so that at least the data that you are getting is valuable.

Given that people wash their hands for a very short period of time, 15 seconds, 30 seconds at the most I think if you are lucky in a healthcare personnel handwash setting, that is the time point that you have to assess because immediately following that wash the provider could go on to do whatever activity they are assigned to. So, neutralization must occur in the test immediately following the wash procedure. This is done in the current method by including a chemical neutralizer in the recovery fluid. This essentially stops the activity of the active

ingredient within a time frame similar to what one sees in washing and rinsing their hands.

In the previous ASTM method and in the TFM method neutralizer was not added until sometime later until the dilution series was created and the samples were taken to the lab. This can occur anywhere from 10, 20, 20 minutes to half an hour after the actual wash procedure.

I don't have the data up here but another study was done. It was presented as a poster at ASTM in, I believe, 2002 that demonstrated with chlorhexidine gluconate that delaying neutralization by approximately 15 minutes increased its apparent efficacy after an initial wash by over 1 log.

So, if we look at the results from the handwashing, the first wash and the final wash, we can see that in the current ASTM method compared to the former ASTM method there is a slight over-expression following the final wash. We would like to see a greater over-expression after the first wash but I think the lab we had do this was

too good and they immediately got to the samples. You can see that following the TFM method you can't even compare the results. So, this makes it incomparable.

The last column is an important point. It is an analytical assessment of how much chlorhexidine gluconate was extracted into the recovery fluid following the wash procedure was measured. While the numbers vary somewhat, the important point here is that all three of those numbers are above the MIC value of chlorhexidine gluconate against Serratia marcescens. Therefore, one has to assume that some activity is going on unless neutralization occurs immediately.

[Slide]

None of these results, however, changes actually in-use effectiveness of the product, and only serves to highlight the importance of determining the appropriate test parameters, as well as maintaining test consistent. Sickbert-Bennet, in a 2004 paper, looked specifically at the ASTM E1174 and the effect that

some test variables, such as product volume and drying time, can have on the effectiveness of alcohol.

The key take-away from this slide is that as alcohol is currently used, and admittedly the N is very small but these results have been repeated in various laboratories around the country. The white bar represents 3 grams of alcohol. To give you a sense of what that is, for those familiar with either the wall dispensers or pumps, that pretty much represents 2 full pumps out of either a wall dispense or a hand dispenser. That is 3 grams.

DR. WOOD: These are two people? Is that what that is?

DR. FISCHLER: Yes. The 7 gram amount would then represent something around 5 pumps from a wall dispenser or a hand pump. You can see that you can achieve a 3-log reduction with the use of alcohol, but the question is are people pumping the alcohol 5 times out of a dispenser, or is the 3 gram amount more realistic of actual practice?

Also to give you a comparison, it takes approximately 30 seconds to a minute on average, and some people are faster and some people are

slower, for 3 grams of alcohol to evaporate from the hands. It can take potentially up to 10 minutes for 7 grams of alcohol to evaporate. So, you can see no one is going to stand around for 10 minutes waiting for the alcohol to evaporate.

[Slide]

So, when the key parameters that can affect data are understood, an evaluation based on the reduction of marker organism contaminating the hand, such as Serratia marcescens or E. coli, is an appropriate way to measure effectiveness. Instead of relying on subject normal flora, these methods control the number of microorganisms on the hand by intentionally inoculating them with a known number of bacteria. In addition, these studies control the dosage, the exposure time to the antimicrobial, as well as other factors.

[Slide]

Our next key points are that surrogate

endpoint testing provides meaningful and appropriate tools to determine the threshold efficacy criteria for topical antimicrobial products, and the published literature represents a body of scientific evidence supporting that the proposed microbial reductions reflect clinical benefit and, importantly, represent current infection control practice.

The SDA/CTFA coalition agrees with the agency that the use of surrogate endpoints to assess clinical effectiveness is a valid mechanism for ensuring that products are efficacious. Surrogate endpoint testing has been used in situations where there is a known benefit, and where standard validated methods have been developed that simulate product use conditions, or where testing and proving a clinical claim would prove to be impractical or unethical.

With surrogate endpoints it is possible to demonstrate a significant incremental benefit from the use of topical antimicrobial products. The SDA/CTFA industry coalition has previously

submitted data on surrogate endpoints that represent clinical effectiveness based on the scientific literature. We agree that while many of the cited studies lack some or the elements found in traditional clinical trials, such as personnel education and training data, incomplete product blinding or specific formulation information, taken as whole, they represent a body of scientific evidence supporting specific microbial reductions and, importantly, represent current infection control practice. The surrogate endpoints that have been proposed were determined from controlled test methods and correlate to a threshold of effectiveness.

[Slide]

Now I am going to focus on each of the healthcare categories, starting with the healthcare personnel handwash. The results from healthcare personnel handwash studies show that a reduction of approximately 1.2 to 2.5 log

5 is achievable

following a single application, and correlate with the literature on benefits of preparations

containing ingredients such as ethanol or triclosan.

I am going to go through these very fast since we have heard about them and we are all aware of the shortcomings that all of the published literature has. But it is important to get some key points from some of these.

[Slide]

This was a study in 1995 that looked at determination of an outbreak of MRSA in a ward through the use of a 0.3 percent triclosan handwash. While not a direct comparison of the product literature, the product used in the study demonstrates a 1.7-log reduction following an initial application and 1.9 following subsequent applications.

[Slide]

A study by Webster in 1994 similarly looked at the introduction of a handwash to eliminate colonization of MRSA cases. A gradual elimination of MRSA was noted and, as a side benefit, fewer antibiotics were found to have been

prescribed--again, not direct cause and effect but another link in the chain.

[Slide]

Hilburn and these next two alcohol studies looked at the use of alcohol as an infection control tool and, again, while not correlating directly, there is strong incidental evidence that the use of the alcohol led to a 36 percent reduction in infection rates over a 10-month period compared to the previous period.

[Slide]

Fendler, in 2002, did a similar study looking at the use of ethanol in a facility compared to regular protocols, and noted a 30 percent reduction in infection rate where hand sanitizer was used.

[Slide]

Dr. Boyce talked at length about Doebbeling so I won't go into that a lot but, actually, what is important to note here is the comparison of alcohol, a product that does not provide either persistence or a cumulative effect

compared to chlorhexidine gluconate that does. Although there were a lot of issues with the study, not the least of which was the use of the product and how much product was used, in a matched pair analysis the authors did find that the difference was directional but statistically significant.

[Slide]

The data supports our previous recommendation that a 1.5 log reduction--and this is based primarily on our review of the alcohol data in amounts as it is used in infection control practice--is sufficient to demonstrate benefit. The necessity for demonstration of persistence or a cumulative effect following several applications of product that is designed for multiple routine applications throughout the day has not been demonstrated. Maybe I should take a moment here to talk a little bit about persistence versus cumulative effect since I think there seems to be a little confusion on the issue.

Persistence is really a demonstration that after a single use typically you have reduced the

resident flora to a certain level and that they do not rebound to a level above what they were when you started. A cumulative effect is very different. It is an application-based phenomenon and looks at what happens after multiple uses of a product rather than what happens over a specific period of time. The definition that Michelle Jackson gave is correct for cumulative effect. It is an apparent reduction in the recovery of organisms. Now, whether that is due to persistence or some other factor hasn't really been well explored. But there is a difference between the two of them. Persistence is time based and cumulative effect is application based.

[Slide]

Surgical scrub products are used by healthcare personnel immediately prior to donning sterile gloves for the performance of invasive procedures to reduce or eliminate transmission of microorganisms from their hands to the patient.

As with healthcare personnel handwashes, surrogate endpoints utilizing a test such as the

ASTM for surgical hand scrub methods have been established for the surgical hand scrubbing in deference to the impracticality of clinical trials to demonstrate reduction of patient infections. In this case, the rate of infection is thought to be very low so any clinical trial would be extremely large and difficult to control. A placebo control would be unethical in this situation so an active control would have to be employed, thus, further decreasing the theoretical differences in infection rates between groups for the study and increasing the sample size. The literature does contain some comparisons between active ingredients, and the coalition has previously presented information that supports initial microbial reductions of 1 log of the resident hand flora, with the flora remaining at or below the initial level, and this is persistence after six hours from baseline. So, in our recommendation we are recommending the demonstration of persistence, not of cumulative effect.

[Slide]

There are two studies--we heard about one of them but I am going to use them for a different purpose, and that is that they both compare alcohol

and, again, we have heard alcohol is a product that does not provide either persistence or a cumulative effect, compared with products that do, either povidone-iodine or chlorhexidine gluconate. In this case, the comparisons are made and I believe are valid in both Parienti and Bryce in that no difference was seen between current practice, which involves the product that did provide a cumulative effect, and a product, alcohol, which did not.

[Slide]

The clinical use of preoperative skin preparations to reduce the incidence of surgical site infections is the most completely tested of the clinical indications contained in the TFM. It has long been considered unethical to even attempt a surgical procedure through intact skin without first cleansing the site, preferably with an antimicrobial formulation.

Given the clinical evidence and the

current standards of care at the time that the 1978 TFM was drafted, the agency acknowledged that the value of the effective skin antisepsis prior to surgery and established surrogate endpoints utilizing the ASTM E-1173 preoperative skin prep method. The coalition suggests that the groin performance criterion of 3 log 10 does not correlate with clinical effectiveness and, in fact, may be unrealistic due to a low bacterial population at that skin site in the general population. The coalition has previously presented information that supports microbial reduction of 2 log 10 on the groin within 10 minutes of use, and again that persistence with no rebound of the resident flora over a 6-hour period, as indicative of clinical benefit.

In one study, and in particular I am going to use this study to also illustrate a point which is that, while it was a comparison of a new skin preparation with a standard 4 percent chlorhexidine gluconate skin prep, two things emerged from the study. One, it was extremely difficult to find a

population that met the baseline criteria set in the TFM. The other point is that the active control product, the 4 percent chlorhexidine gluconate, did not achieve the log reduction required from the TFM. It achieved a 2.5 log reduction following 10 minutes of application.

[Slide]

One of the performance criteria, addressed under patient preoperative skin preparation in the TFM, is the pre-injection skin preparation performance criterion of 1-log reduction of skin flora within 30 seconds of use. The coalition agrees that this is a suitable surrogate endpoint for clinical efficacy for this indication.

Clinical trials for this indication would be possible but impractical. As with the previous indications, injection site infections are a rare occurrence and would require a multiple-day follow-up period to assess the infection rate. Therefore, the surrogate endpoint for these studies is a reasonable alternative.

[Slide]

In conclusion, we would like to emphasize the following key point. The efficacy criteria of healthcare antiseptic drug products should be

appropriately set to reflect the performance of currently recognized effective products. Thank you. Now I would like to introduce Jim Bowman who will address some issues on statistics.

Statistical Issues in Study Design

DR. BOWMAN: Good afternoon.

[Slide]

I am Jim Bowman, technical director, biostatistician at Hill Top Research. I too represent the CTFA/SDA coalition. I have been asked to summarize the statistical issue at hand.

[Slide]

Log reduction criteria has historically been based on point estimates with no set requirements for sample size. It is understood that variability needs to be considered, and there are several ways to take that into account.

[Slide]

Here are two examples that come from other

OTC monographs. From the sunscreen monograph, a mean value is calculated and then the standard error is used to calculate the SPF value for product labeling.

From the antiperspirant monograph, in order to label a product as an antiperspirant the tested mean value must be statistically significantly greater than 20 percent sweat reduction.

[Slide]

Our objective is to obtain a mean value greater than or equal to a certain log reduction. With point estimates manufacturers have historically conducted studies with sample sizes they deemed appropriate, and submitted data to the FDA. With statistical criteria being utilized, i.e., statistically greater than a specific number, appropriate sample sizes are a function of the variability of the data.

[Slide]

We have conducted data reviews and statistical simulations using data from hands and

looking into the variability. This review consisted of data from 13 studies conducted with an active material, and simulations were conducted to better understand that variability. Our conclusion was that if statistical criteria are to be utilized, then lower criteria will be necessary to achieve the same level of efficacy based on our data review.

[Slide]

For an example we can look at the antiperspirant monograph. The OTC antiperspirant monograph requires statistically significantly greater than 20 percent reduction. However, this requires point estimates of sweat reduction to be greater than 25 percent to 30 percent reduction in order to achieve the level of benefit mandated.

[Slide]

Historically, the FDA and industry have relied on point estimates. All recommendations from the coalition have been based on point estimates. However, if statistical significance is required, then lower log reduction criteria are

necessary to achieve the same level of efficacy based on our data review. We would like to work with the FDA on setting these criteria for specific indications at specific time points. Thank you for your time. George will now summarize.

DR. WOOD: Just before you step down, can you put up slide 23 from the last talk? I have a statistical question on it. This has been offered to remove one of the criteria. So, what was the sample size in this study, and what was the size of the difference that you could exclude, and at what power?

DR. FISCHLER: I have to refer to the paper for that.

DR. WOOD: Well, this is being offered as one of the key pieces of evidence.

DR. BOYCE: I am not sure, but I think they are in that range of 1,500 to 1,800 patients in both arms of the study so there were over 3,000 patients that underwent surgery during the trial.

DR. WOOD: And what was the size of the difference?

DR. BOYCE: The difference between the two arms was about 0.04 percent, in other words, no significant difference and it was considered to be

an equivalence trial.

DR. WOOD: So, it was set up with some sort of power calculation in advance?

DR. BOYCE: Yes, I believe so.

DR. WOOD: But we don't know what that was?

DR. BOYCE: I have the reference here.

DR. FISCHLER: And I have the paper.

DR. WOOD: The second part was it was possible then to do a clinical study. So, you feel that this was an adequately powered study to show a non-inferiority outcome and it only needed 3,000 patients.

DR. BOYCE: I think that was the conclusion that the authors arrived at.

DR. OSBORNE: Dr. Wood, just to review the exact data from that study, there were 4,287 patients, divided roughly equally into the three groups, the alcohol hand rub, the PI and the CHG.

The surgical site infection rate was 2.44 percent for the alcohol versus 2.48 percent for the combined group of PI and CHG. What more can I give you?

DR. WOOD: That is fine.

DR. OSBORNE: That is where the 0.04 came from that Dr. Boyce mentioned.

DR. WOOD: And Tom can calculate that on the back of an envelope, and probably already has. Frank?

DR. DAVIDOFF: What was the confidence interval of the difference?

DR. WOOD: I don't know.

DR. DAVIDOFF: Isn't that the key question?

DR. WOOD: Do we know that?

DR. BOYCE: I don't know the confidence interval.

DR. BLASCHKE: It had to be pretty small if the difference was 0.04 between the products and they were statistically significant.

DR. DAVIDOFF: That is how you can talk

about meaningful exclusion of differences. Without that, it is real tough to do that.

DR. WOOD: All right, let's let him finish.

DR. FISCHLER: I will be very quick in rapping up. In summary these are our points.

[Slide]

Definitive prospective, and controlled clinical trials are not practical in measuring the prophylactic benefit of antimicrobial products. Again, I think we have to look at these as three different types of antimicrobial products, the healthcare personnel handwashes, the surgical scrubs and the preoperative preps.

I am just going to make one point which is if you look at those three and you start with the preop prep--I forget who said this in their presentation, but in a preop prep the patient represents their own control. So, you have basically the smallest denomination. If you look at surgical scrubs you are not looking at the benefit being derived from the surgeon, but it is

essentially a one-on-one calculation, the surgeon and the patient. When you move to healthcare personnel handwashes, you are now trying to look at what is the benefit derived to a general population from another population that has used the product? So, in equivalence it is asking the question what benefit do the members of the committee seated at the table derive from the people in the audience washing their hands?

Standardized, defined, and peer-reviewed test methodology, such as ASTM methods, encourages reliability, reproducibility and comparability of results. Surrogate endpoint testing provides an appropriate tool to determine threshold efficacy criteria. The published literature, with all its shortcomings, supports that the proposed surrogate endpoints represent clinical benefit. Finally, the efficacy criteria should reflect the performance of recognized effective products. And, I will be happy to answer any questions.

DR. WOOD: Questions from the committee for the last two presenters? Ralph?

DR. D'AGOSTINO: If I understood the FDA presentation, the literature was full of studies that were inconclusive, and we have heard some

fairly definitive statements, I thought, with this presentation. Could the FDA respond to that?

DR. WOOD: Well, I am not sure that I agree. Some of them had two subjects in them.

DR. D'AGOSTINO: I would have presumed that the response to my question is going to be that it is a rosier picture than what is real.

DR. WOOD: Right.

DR. OSBORNE: If there is a request about a comment on a specific study, I could make a comment on that specific study.

DR. D'AGOSTINO: Well, the one where the sample size was 4,000 and you gave the numbers, was that a well-designed study, well executed?

DR. POWERS: What I was trying to answer about that previously was that we struggled greatly with how to interpret non-inferiority trials in this setting. To look at that study, regardless of how many patients it has in it, and determine that

two things are not inferior to each other means one of two things: Either both products are effective in doing something or neither product is doing anything. The problem is that without the ability to determine what the magnitude of benefit over whatever you want to specify as the control is in that over nothing, it is very difficult for us to interpret what no difference actually means in this setting.

So, what we really want to look for is trials which showed some kind of a difference, and that was very difficult to find. Then, when you did look at those trials, many of them actually had flaws in them in terms of there was no concurrent control group or other things that made it very difficult.

So, we did not just look for a p value at the end. We asked the question of how did you get to that p value, and that really had a lot--the buzz word "evidence" has gotten thrown around a lot here today, and just because you have lots of studies, does that really mean that that is

evidence or not? That is one of the things we struggled with in a 1,000-paper review.

DR. D'AGOSTINO: It does go back somewhat to the discussion we had half an hour ago about vehicle control and positive control and trying to interpret in that setting, I agree.

DR. WOOD: Other questions for the speakers? Tom? DR. FLEMING: I come up just crudely with about half a percent, just to go back to this slide 23 where you had 2.44 and 2.48 and you could rule out a 0.5 percent difference but what does that mean? If you are essentially the same and you can rule out not more than a 0.5 percent difference, are you the same effective or are you the same ineffective?

There are several things on your slide, this last slide. The last point says, and it is reworded from an earlier conclusion slide where you had said efficacy criteria should be set to reflect the performance of concurrently recognized effective products. What is the effect of currently recognized effective products? If I know

that a currently recognized, effective product provides a 50 percent reduction in infection risk and I have a lot of studies that allow me to understand that I need to achieve a 2.5 log reduction to achieve that, and the relationship is if I give up half a log reduction that I am giving up 10-20 percent protection on infection risk I am buying into your last statement. Tell me how I can, in fact, address, based on currently available data, how much efficacy--or I would call it how much biologic activity I have to achieve in reducing log reduction in bacterial load to achieve clinically meaningful benefit on infection risk.

DR. FISCHLER: I guess not to give you a smart answer, but I think that is what we are here to try and determine. I think we are struggling as an industry with the same issues that clinicians have been struggling with, which is that we are operating under a regulatory framework and when we look at infection control practice today, specifically highlighting the fact that alcohol hand rubs have become a key part of infection

control, and looking at how alcohol hand rubs are used in infection control and what does that translate to in a surrogate endpoint test--and the determination of whether or not surrogate endpoint is appropriate or whether or not the test is appropriate we will set aside for a moment--but looking at that, if we admittedly go back to the Sickbert-Bennet with an N of 2 but companies do have internal data that did repeat that study. There is probably data on several hundred subjects doing that exact same study. Most people put 2-3 grams of alcohol on their hands at most and what got a log reduction in a standardized test to come up with for that 2-3 gtam amount of alcohol.

The issue that we are all struggling with here is while that is all well and good, how does that translate to a clinical benefit? I think we have heard from pretty much everyone here that no one can definitively say that any of these log reductions translates to a clinical benefit in terms of the way clinical trials are assessed. So, in my own poor way I guess I am saying what we are

trying to do is not lower the efficacy standard but match. Over probably 20 years of infection control practice people have been washing their hands in hospitals and using antiseptic products for over 20 years. Dr. Larson stated it when she said the horse has left the barn. We are trying to look back at over 30 years of data and saying what is going on and what is happening.

So, what we are looking at is current practice, and if current practice is acceptable, and I can't answer that, only clinicians can answer--if compliance is not an issue, if infection control practice as it is currently performed today meets the standards of care, then for the products that are being used we should analyze what surrogate endpoint test results they achieve in whatever standardized test we come up with so that we don't set criteria that essentially will eliminate the products that are currently being used for infection control. That is a really long-winded answer. I don't know if I got to the heart of your question.

DR. FLEMING: It is a long-winded answer. I guess my short interpretation of the answer is we could, in fact, justify using surrogates if, in

fact, we had evidence that allowed us to know what the actual efficacy is of currently used products or efficacy on prevention of infection, and where the data were also allowing us to understand how the influence on bacterial load was causally leading to what the association is with the reduction in infection. We lack that evidence--

DR. FISCHLER: Correct.

DR. FLEMING: --therefore, we lack the ability to draw that conclusion. You went on to say, well, then we will ask clinicians whether what we currently have in the real world is adequate. Dr. Pearson, in her presentation, said we have 2-5 percent infection rates with surgical site infections. Is that adequate? I think we would all say it is better than 8 percent; it is not nearly as good as 0-1 percent. Now the question is how do we achieve 0-1 percent? What are the interventions that are out there that are more

effective than others? How do we determine how maximally to use them?

Let me just close by saying you made the point earlier on that we have a complicated situation, and that complicated situation is multidimensional involving immunological host factors; involving test subjects versus populations; involving compliance. And, for this reason, clinical trials are not appropriate. I can look at a lot of other areas. An area where I am involved in my own research, which is looking at vaginal microbicides as a way to prevent heterosexual transmission of HIV where, clearly, all of these issues are relevant and many of us are embarking on major clinical trials to answer the question as to how these interventions affect transmission rates. So, this isn't a unique challenge.

DR. FISCHLER: I guess I would go back to the regulatory framework within which we have been operating for the past several years, which is the world of surrogate endpoints from the FDA's

perspective. I think our key challenge, and I think it is reflected in the questions that the committee is being asked is, is that the world we should be in? Is that appropriate? Should we be moving somewhere else? And, how do we deal with the situation moving forward because there has to be common ground somewhere?

DR. WOOD: I think what Tom was also asking you is this, you are here today proposing a reduction in the surrogate standard, rightly or wrongly and I am not arguing with that right now. What I think the committee would like to hear is what is your estimate of the clinical outcome of that reduction in the surrogate standard and point us to where we would look to see the evidence to support that.

DR. FISCHLER: I don't know that you would see a reduction because is practice going to change? Practice as it exists now will not meet that standard that is set. So, I guess it is a question of if you change what is printed on the page, does that change infection control outcome?

Or, as we are suggesting, do you match products? Do you find a test that everyone can agree on that adequately measures whatever outcome you are trying to measure and then determine what the number should be?

We feel that the number as published in the monograph has a number of flaws, the cumulative effect for healthcare personnel handwashes among other things. We feel that the number, the 1.5 log reduction reflective of alcohol under use conditions, is reflective of current practice.

DR. WOOD: That would be terrific if we had a zero percent infection rate, but we don't have a zero percent infection rate and, given that, what has led you to believe that we are currently in the ideal Nirvana?

DR. FISCHLER: I guess I would ask the question is zero a number in a biological system? But besides that, I guess I can't answer the question of is 2 percent to 5 percent acceptable. Certainly, the lowest number that is possibly achievable is the goal. But setting a standard for

current products--I guess that is what the committee has to decide, changing the standard so that products that are currently used are no longer available because they do not meet the standard--will that increase or decrease the public health?

DR. WOOD: Any other questions?

DR. BRADLEY: Just a clarification. The TFM from 1994 sets some criteria and guidelines. Yet it seems in this discussion that the alcohol-based solutions don't meet the TFM guidelines, yet they are being used and recommended. So is it true that the current TFM guidelines aren't being enforced with these products? If that is true, then the industry is asking for a further reduction even though we have a standard that is not yet being enforced. If we, as a committee at the end of the day, feel that the TFM standards should be enforced, then we should raise the bar from where we are right now.

DR. LUMPKINS: Basically, because the OTC monograph process is a public rule-making and a

multi-stage process, what the agency has decided as a matter of policy is that we don't enforce proposals. So, right now, there is not a requirement for anybody to comply with the TFM.

What the discussion today is about is what do we finalize and what, at the end of the day, will everybody need to comply with. That is what you need to worry about.

DR. WOOD: Jan?

DR. PATTERSON: I just wanted to comment back on the Parienti study, the surgical site infections. These are two antiseptics compared to each other so there is not a control, which I think was the issue. But I don't think that an IRB committee would approve the study of surgical scrubs that didn't involve an antiseptic. And I, personally, wouldn't want to be a subject in one in which my surgeon might not have used an antiseptic.

I think there are some practical considerations like that. Even as was mentioned this morning, you might be able to do a plain soap versus an antiseptic for routine patients on the

ward but now we have a federal guideline from CDC that says we should be using antiseptics and also an accrediting agency advised by federal agencies tells us this can be monitored. So I think it is very difficult to talk about comparing antiseptics to non-antiseptics.

DR. WOOD: Dr. Patten?

DR. PATTEN: I have a question for the FDA. If the requirements that you are proposing in your TFM were to be finalized, what sort of a time frame would be built in to allow the industry to respond?

DR. LUMPKINS: Once a final rule is published there is usually a one-year period for implementation. However, I have to be honest with everyone involved. In the monographs that we have developed that have required final formulation testing, in reality there have been a number of stays of the final rule to allow industry time to make adjustments.

DR. WOOD: And some have never got to final, right? Let's be honest here.

DR. LUMPKINS: Hopefully, we will fix that.

DR. WOOD: Right, but I mean there is a

lot out there that has never got to final. So, it is not a door that has to be closed. Dr. Larson?

DR. LARSON: Despite all the difficulties of answering all the questions we have to answer, and I don't know the answers either, I just want to point out that this has been a tentative final monograph, first in '78 and now in '94 so for decades it has been tentative. In some ways, patient safety is more at risk by not finalizing something because now products can be on the market and there is no regulatory agency that is overseeing them by force of law. So, even if we don't all agree on what it should be, I would hope that it wouldn't stay tentative until after I die, for example, or until after my career is done because I have been waiting for 30 years for a final ruling of some sort and, in the meantime, the good industries want to do it right and they want to follow the rules, and they ultimately, I am

sure, have the same goal we all do which is to reduce infections. But right now it is possible for industry to be out there, selling something that is inferior, because there are no rules.

DR. WOOD: Mary?

DR. TINETTI: I think we do need to discuss separately surgical scrubs from the handwashes. I agree with that. I would think it would be very difficult to do anything in the surgical scrub at this point. But for the handwash, I mean what we are hearing today is that there is no evidence linking the standards that are in the TFM with the clinical outcome that we are interested in. We are hearing that guidelines exist in JCAHO but guidelines were developed in the absence of evidence and to now use those guidelines that were forced because there was lack of evidence as a reason for not procuring evidence seems to me a road I certainly would not like to see healthcare go down.

Certainly, this may be the final opportunity for us to preclude that from happening

and I think there are alternatives to study it. Yes, it is difficult but a lot of us do research that is difficult. Difficult is not a reason to preclude it from happening. Yes, it is going to be expensive but these are marketed because they say they do improve the clinical outcomes and the fact that, yes, we treat the healthcare providers to help the patients, that is what these are marketed to do so it seems to me that these studies in that area are feasible. I think it will be setting healthcare back to finalize it when there is really complete lack of evidence.

DR. WOOD: Any other comments? If not, let's take a quick break and come back at 3:10 and we will start the final discussion and deal with the questions. So, 3:10.

[Brief recess]

Committee Discussion

DR. WOOD: To summarize what I think we have heard so far as we begin the discussion, I think what we heard--I tried to jot down some notes here--we heard that there are no adequately

designed or powered studies to demonstrate the clinical effectiveness of these topical antiseptics. Given that, therefore, it is not surprising that there are no adequately designed and powered studies that demonstrate the robustness of any particular surrogate in predicting the clinical effectiveness of these agents.

As I think Susan or somebody said, the standards are arbitrary but steeped in history, and industry clearly believes the current products are clinically effective but industry wants to lower the bar for the surrogates because they have products that can't meet these standards. Industry has no evidence that lowering the standards for the surrogates won't impair effectiveness and result in patients being at increased risk for infections, again not surprising given the current lack of clinical correlates for the surrogates in the first place.

So, I guess I don't see how, in the absence of data, we can possibly endorse lowering a standard for which we have no evidence that it is

clinically relevant and when we can't determine what would be a safe reduction in that surrogate in the first place.

Finally, I don't see how industry, or anyone else for that matter, can argue that if they believe the current products work, whatever that means, that products that work less well, again whatever that means, can possibly be approved without someone going out and doing a study to determine the clinical consequences of that reduction in effectiveness.

So, it occurred to me that a way out of this dilemma, Susan, was to ask you this question. We are working around this sort of mish-mash of the historical precedents, but supposing somebody were to go out and do a study where they demonstrated that their product reduced bacteremia--all the things that we have heard are impossible to do, but supposing somebody did it, would you approve a study and give them that as an indication?

DR. JOHNSON: There are a couple of issues here. Let me put the clinical one aside for just a

second and talk about the regulatory. Within the monograph--

DR. WOOD: No, no, I am not talking about the monograph. I am saying forget the monograph for a minute. Somebody goes out and does a study in which they demonstrate that X, Y, Z handwash, or whichever indication it is, reduces bacteremia in patients, or some other hard endpoint and you can pick whichever one you like, would they get an indication for that and would they be allowed to promote on that basis?

DR. JOHNSON: We have been asked various permutations of that question many times from NDA sponsors, and we have always supported that under an NDA were they to come up with a clinical design and conduct a trial that showed that sort of effect, we would label them accordingly.

DR. WOOD: So, one of the things this committee could do in addition to answering the questions is to come up with that as a proposal, which would get us out of Dr. Larson's very reasonable point that she wants to live long enough

to see this finalized. Essentially, it seems to me there are two tracks we can take. One is to promote the rational adoption of a regular process, which would be to find a clinical endpoint and do that, or not if you can't do it, and the other is to proceed down the current track.

The attraction of the former, which is the clinical endpoint, is that clearly any sponsor who does that and comes out with such an endpoint trumps everybody who is unable now, they say, to do that, which would obviously be a very compelling argument both in the marketplace and hospitals who purchase these things and, I guess, the JCAHO. So, that would be a reasonable approach from the agency's point of view. Is that right? All right. In that case, let's move on to discussion and who would like to comment first? Yes?

DR. LEGGETT: I have a question for the FDA. Suppose we come up with a final monograph, what happens to the products that are on the market already?

As a corollary to that, I would like to

mention this Sickbert-Bennet paper that we got just this past week, in AJAIC this month I believe. In table 3 they looked at the log reductions of Serratia marcescens in the hand hygiene agents, albeit this is with that 10-second wash because they document elsewhere in the paper that the median time of washing hands is 11.6 seconds, or something like that. In agent A, which is the 60 percent alcohol, the first wash only had a reduction of 1.15 logs. So, even by the industry's standards this would not fly. At episode 10 the alcohol actually had a negative trend; it was less efficacious after 10 washes, and they had some theories in the paper. So, with those two questions, what does the FDA do if you get a final monograph?

DR. JOHNSON: One of the things that I would like to do is ask Colleen Rogers to address the information that she found in doing the literature search on the handwashes. But as a general response to your question, the products can be formulated, as far as we have seen in the

literature, to be able to accomplish what we have proposed in the tentative final monograph. This alert that is being sounded that in general the products are failing to meet this standard is not what we have observed in general in NDA submissions. We don't see data submitted routinely under the monograph because that is not the way the process works; it is dissimilar to the NDA in that regard and it is driven by the literature. So, we are not seeing the same level of current studies coming in under the monograph prospectus. But in reviewing the NDA data, which obviously we can't present to you, we are seeing that this is not an across the board uniform problem.

If I could ask Colleen, there is a difference between the immediate acting alcohol products and the leave-on products, and the difference is in formulation and she might want to comment a little bit more on the data that we found.

DR. ROGERS: In reference to what Dr. Johnson was just saying, in looking through the

literature most of the alcohol-based products are leave-on products and they are not rinsed off the skin. Compared to what was presented in the most recent Sickbert-Bennet paper, those products, for one, were used for a very short time, 10 seconds, and most of the other studies that I looked at used a longer time period for contact with the skin.

Also, if I remember correctly, in the recent paper, the Sickbert-Bennet paper, they also rinsed after using an alcohol product, which is not normally done with an alcohol leave-on product, and that may have affected the results in that most recent paper.

DR. JOHNSON: I would just add also that one of the things that we are very interested in resolving for the final monograph is to be sure that the test methods reflect the intended labeling. Some of the variability in the responses from the current test methods are because we are not clearly using the intended labeling activities to do the wash. So, that is where you see these variations.

Getting back to the point that you have been trying to make, the fact that people only wash their hands for 10 seconds is not a good reason to

label products or test them that way.

DR. WOOD: Right. So, in response to Dr. Leggett's question, I guess you are saying that in looking at the totality of the products that have been approved, there is not going to be no products there tomorrow, which I think is what you are asking. Is that right?

DR. LEGGETT: Yes.

DR. WOOD: All right--

DR. LEGGETT: Because the purpose is to wash hands and if we find that more people are washing their hands--I don't care if it is for 10 seconds but if it is every minute, every door they go in and out of, that is what our goal is eventually.

DR. WOOD: Right. Although, interestingly, we have no data to support it, it sounds like.

DR. LEGGETT: Right.

DR. WOOD: Any other questions? Yes?

DR. BRADLEY: I would like to go back to the question of cumulative effect, not just over a day but over two to five days. In the surgical hand scrub requirements, it appears that there is a day-2 wash and a day-5 wash, and the day-2 wash is

wash 2, and the day-5 wash is wash 11. Certainly I can clinically understand why you would need several hours of cumulative effect, but to have criteria where you still need effect at day 5 I don't understand fully. Do you know the rationale behind that?

DR. LUMPKINS: Like I said, a lot of this has been lost to time. There may be people in the audience who developed these methods who might be able to speak clearly to your question. It is intended to mimic actual use where handwashes get used numerous times during the day.

DR. WOOD: Dr. Bradley, did that satisfy you? Mike?

DR. ALFANO: Thank you, Mr. Chairman. This will be probably a longer comment than I will

make in the rest of the day so maybe you will indulge me for a second or two. You know, when this process started, Alastair, you and I had hair.

DR. WOOD: Long hair probably!

DR. ALFANO: I actually see that as not necessarily an argument to speed up but as an argument to be cautious in the absence of data, as we have seen here today. So, my comment revolves around that and the way we are looking at data these days. So, I am troubled. I applaud the agency for getting us here and trying to get at the clinical data sets that are desperately needed. I am troubled by the fact that in over 1,000 studies not a single one was deemed worthy of presentation as a model as to how these things might be done in the future. So, take money off the table for a minute and I will come back and comment about money.

I would worry that the industry would be able to design trials that would meet these new higher standards, and you don't only see it with regulatory agencies; you see it in academia as

well. While I applaud the concept of evidence-based reviews, there tends to be an intellectual elitism around them, that, you know, no one can do these trials, not even me, and the people who do these reviews tend to sit on the mountain top and cast aspersions at the people who are trying to get some clinical data done.

I am going to give you a practical example of something I lived through because today this has been deja vu for me. I had the opportunity to chair, about two and a half years ago, the NIH Consensus Conference on Dental Caries. It was the first time NIH started feeding evidence-based reviews into the panels that do these reviews, clearly the first such review.

The problem was that the evidence-based reviews selected a standard for measuring dental caries that is virtually unattainable. What they said was that looking at radiographs of tooth decay and watching the dentist use the pick, as people like to call it, are really only surrogate markers and the only way we know a tooth has been affected

is if we extract that tooth and section it. So, they dismissed all of the other studies that didn't extract teeth. So, you know, I have this vision of a parent signing the consent form and at the end, "we will extract all your child's teeth."

[Laughter]

I am not making this up. There are people who can validate this for me. Curiously, there are some studies that were done on extracted teeth and they were deciduous teeth that exfoliated naturally and were collected at the end of the study. So, my great fear as chair of this conference was, you know, a front-page story in The Times, "panel declares fluoride ineffective" because we essentially threw out everything. Thankfully, that panel recognized that there was a preponderance of data and that, while there wasn't a definitive link to the value of these surrogate endpoints, radiographs in this case, it was good enough not to come out all the way on the downside.

A second fundamental point is the size of the industry. I have heard some panelists intimate

that, you know, certainly we have seen the pharmaceutical industry doing studies that are $30 million, $40 million, large heart trials for example. Clearly, this must be a market that we are talking about today that is in the billions of dollars. As the industry liaison, I asked for that data and the latest four quarters, so full year data, is that it is a $237 million market, and it is described almost as a commodity. To translate it for the people who haven't spent time in business, that means very low profit. As opposed to a Lipitor which is 8 or 9 billion and a very high profit product.

So, the idea that the industry is sort of stingily applying funds to this problem is probably inappropriate. You know, maybe the profit margin here is 10 percent, 8 percent, in this category. So, you are talking about across all companies $15-18 million of additional revenue, $20 million maybe, that could be spent. And, I think we need to frame our discussion along those lines because the concern then becomes, well, we can't do that;

we can't do that study; we are just exiting the market. We have seen it happen. We have seen the problems this country faces today because vaccine manufacturers have exited the market, not because of pressure from the FDA but because of pressure from the trial lawyers because any child that is born today with a defect--someone has to pay--it must be the physician; it must be the vitamins the mother was on; it must be something. It is not me; it is not my genes. Someone has to pay. So, companies just said we are exiting; we can't make any money in this arena.

So do we stop? No, we don't stop. I am not proposing westop. I have a good possibility personally of going under the knife based on the odds presented here today and I would certainly like to know that whatever is being used is going to work. I am pleased to see that, you know, Columbia has been funded in the nursing program through the Road Map because I think the NIH Clinical Road Map is clearly an area that could provide funds to do these larger scale trials to

try to benchmark a surrogate endpoint, not so much to look for a specific product going forward.

I am concerned that one of the pieces of data I saw would eliminate NDA products. Chlorhexidine didn't pass, at least in the study that was shown by Dr. Fischler--it didn't pass; it didn't come close to passing the newest tentative final monograph. So, what does that mean in terms of availability of products?

I guess I will conclude by sort of drawing on something Dr. Powers said only using it a different way, and that is unintended harm. We could potentially, if we are not careful, do unintended harm by removing products that may have a benefit although, admittedly, we haven't demonstrated that benefit, and I wouldn't really want to be a part of that approach. Somehow or other, it is calling almost for a starting over type of philosophy in which people of sound mind and good intentions get together and determine in advance what would be acceptable to validate these surrogates and move on from there.

DR. WOOD: There are lots of approved drugs which have also failed. You know, actual pharmaceuticals that have also failed in clinical

trials and we think they are effective. That doesn't mean that showing a single trial means that antidepressants don't work, for instance. It is a good example where frequently trials fail to demonstrate efficacy. So, I don't think that should make you too pessimistic just because somebody can find a trial that shows something doesn't pass the test.

DR. ALFANO: I think that is a fair point. One other comment, by the way, about evidence-based reviews. I think there is something missing on the high ground in evidence-based reviews so when you do the A category trials it doesn't allow for an FDA reviewed and audited trial to get a higher level. So, when evidence-based approaches became the rage I said to myself, well, wait a minute, how can FDA be approving new drugs with two trials, and sometimes only one trial? I realized there is a difference, and that is that for those trials every

piece of paper, every data point is sent into the agency and frequently the sites are audited. So, there is another flaw in the way we rank evidence-based assessments that I really think somebody should look at.

DR. WOOD: That was actually discussed in a New York Times article recently. Frank?

DR. DAVIDOFF: Yes, I would like to pick up on the comments that you just made because I think in hearing how much the agency, understandably, is pushing for a certain kind of evidence--randomized, controlled, and so on, I think what that tends to lose sight of is that then, in a sense, all of us have become what I would describe as prisoners of frequentist statistical methods.

I would like to suggest seriously that the agency consider undertaking a formal Bayesian process. I am not a Bayesian statistician; I am not a statistician but I think I understand enough about the difference between frequentist and Bayesian statistics to understand that the

intrinsic logic of frequentist statistics is actually weak and that Bayesian statistics has its own limitations but it gets around that fundamental weakness of frequentist statistical methods. I hope the statisticians here will not take me out afterwards and beat me up.

I think one of the big problems with frequentist statistics is that essentially that approach forces you to make conclusions on the basis of each individual study or, in effect, all prior information is ignored in coming to a conclusion about the results of each individual study. It seems to me that is an enormous waste of information. I mean, we have sat here all day and spent hours before coming here to be saturated with very large amounts of important information that is characterized as kind of background, and that is the background on which you will sort of then interpret the results of one or another individual paper but the background isn't taken into account in any formal way in interpreting any particular study.

Whereas, Bayesian approaches to making decisions basically consider all the prior evidence from all different sources and they are all

integrated into an initial degree of confidence in the validity of some phenomenon in the real world. The problem with that, of course, is that that initial sort of conglomerate degree of confidence is a subjective judgment and I guess that is the big drawback for a Bayesian type of approach.

So, while a frequentist approach avoids the subjective element, it does have its own drawbacks. But I think there are ways to sort of get at the problem of subjective limitation of Bayesian priors. One approach is to combine the initial or prior subjective judgments of degree of confidence across a group of experts, for example the people in this room. In a way, that is the process that is part of what I think we have been hearing going on today.

Once that is done, then the additional information from each individual piece of evidence that is considered at least partly credible can

then be used to modify that initial degree of confidence using essentially a likelihood ratio as the modifier, the so-called Bayes factor as Steve Goodman calls it.

I would suggest that that approach might be a somewhat more formalized way of getting off the dime than just sort of saying, well, we don't have enough evidence and the reason we are saying that is because the evidence, if it isn't perfect, essentially is being rejected. I think that is a problem, albeit there are problems the other way, of course, that is, you don't want to go taking a thousand papers that are weak and adding them up and saying, well, that adds up to strength, which is not I think what good Bayesian reasoning does anyhow.

I would like to make one other suggestion looking ahead, and that is that there are other ways to gather data in a rigorous way that don't involve the usual p value testing of aggregated data, and that is essentially using time series data in the process known as statistical process

control. There are rigorous statistical criteria that can be used that actually are very powerful in examining data spread out over time which give you the time history rather than a collapsed or snapshot view. I would suggest that if data of that kind had been collected and used in some of the studies that we are being presented with, I think there might actually have been much more compelling evidence on efficacy or lack of it than has been made available just through these kind of snapshot, cross-sectional statistical analyses.

DR. WOOD: Dr. D'Agostino? No? Anyone else? Yes?

DR. BRADLEY: Just another quick regulatory question, I am sorry. How flexible is the final monograph going to be for allowing people to use different ways to use topical antiseptics? So, if someone wanted to spray the wound with an antibiotic-containing solution after opening every 15 or 30 minutes, if it is not in the monograph is it not considered? Or, is that another agency? Or, do you have flexibility in the final product?

DR. WOOD: Well, the monograph wouldn't consider the use of the product. That would be the practice of medicine. Right?

DR. JOHNSON: Well, the purpose of the monograph was to corral all the products, the active ingredients that were on the market pre-'72. The way the formulations have been modified over time--we make decisions on a case-by-case basis really about how the translation of those active ingredients into new formulations does or does not fit in the monograph. We can talk about some precedents but we would actually have to make an active decision about something that was very different, that would end up having a very different indication.

We are actually considering discussing the alcohol leave-ons in almost a separate category for how we would actually formulate labeling for those. So, it is a little difficult to project. What a creative idea though. I think you could probably sell that here today. But I couldn't say how we would actually address that in the monograph.

DR. WOOD: Is there a way for us to think about the monograph sort of proposition here, that this is the equivalent of bioavailability comparisons for essentially topical products, or something like that? Because these products might contain something that removed the efficacy of the

antiseptic--we obviously don't want to just measure concentration so we are measuring the equivalent of a bioavailability comparison for a generic, or something like that. Is that reasonable?

DR. POWERS: That is actually a good analogy in terms of suppose you wanted to test a new formulation of a particular drug that had already been proven effective in the treatment of community-acquired pneumonia--

DR. WOOD: Right.

DR. POWERS: --and all you wanted to do is say, okay, we are going to change it from a tablet to a suspension but you are looking at the same active ingredient. But, John, your question is we don't want to study it for pneumonia anymore, we want to study it for meningitis; we want to look

for a different indication. So, as John knows very well, everything at the FDA starts with the claim you want to make, and the monograph has very specific claims associated with it. As Michelle Jackson presented, there are three of them in that monograph. If you want to deviate from that and look at some new use such as putting something in here to prevent catheter-related bloodstream infections, that gets shifted over to the NDA process because that is not covered within the monograph.

DR. WOOD: So, for the monographs we are talking about we are really talking about trying to create a comparison between similar products and demonstrate they still have the same in vitro effect. Is that a way to sort of formulate the issues? Dr. Larson?

DR. LARSON: I think the fluoride analogy was a good one and I would just like to clarify again the difference between the clinical evidence that is out there and the kind of evidence that FDA needs and this panel needs. The clinical research

asks the question, given the products that are available, what is the evidence of effectiveness in the clinical setting, relative effectiveness or whatever. The FDA's rule is written to say what is the level of safety and efficacy that we need before we allow a product to be on the market. It is very different.

But I did just want to clarify one thing. The clinical practice guidelines are based on evidence. It is just a different kind of evidence. And the CDC, the two or three years that they spent developing this guideline--it is a different kind of evidence than one would use for making rules. So, it is not that there isn't evidence out there; it is just that it is asking very different questions and I don't think that the clinical evidence that is out there relates to what the panel needs to decide about the log differences, etc.

Your point earlier this morning, Frank, about is a log reduction even relevant, and do we need other kinds of statistical modeling or do we

need to set a baseline--a 1- or 2-log reduction from 7 logs is quite different than a 1-or 2-log reduction from 4 logs.

DR. WOOD: Dr. Snodgrass?

DR. SNODGRASS: Well, I think we need clinical trials on some level. I guess the question is, within the limits of how the FDA can operate, can you put some wording that clinical trials are strongly encouraged? I don't know if what the issues are in incorporating some kind of language like that.

The other issue, and I would just add to what has already been brought up, is if you have a specific, for example, bacteremia, that is a step. That is a really good step.

DR. JOHNSON: I guess there are two points in there. One is the variability available to us in the monograph process. The variability in how to address specific questions is designed in the monograph to be very limited so we could encourage clinical trials under an NDA and if folks wanted to default to using surrogates, if that was still an

acceptable method, that would be something that we could discuss with them in their development programs. But under the monograph we don't really have the flexibility to say either/or, not in such a wide variation. I am sorry, I lost the other point.

DR. WOOD: Clinical trials I think.

DR. SNODGRASS: Yes, clinical trials in the specific of choosing some endpoint that can be measured, that is achievable, like bacteremia as an example.

DR. JOHNSON: Right. Anything that would significantly differ from the monograph indications--and Dr. Rosebraugh has pointed out to me there is a process that is called an NDA deviation which is similar to a 505(b)(2) and relies on the monograph to some extent, largely for the safety component, and is a limited development program that might be applicable. Again, it would go back to Dr. Bradley's question about how different is the formulation. At some point they become diverse enough so that the regulatory

processes can't lean on one another.

DR. WOOD: But we shouldn't lose sight of the huge advantage a product would have in the marketplace if they came in with some sort of endpoint that was clinically relevant. While I take Mike's point about the size of the market for an individual company, a company that came in with a product that had that kind of block-buster effect would make huge amounts of money. I mean, it would be hard for any hospital to use any other product in the face of that setting. I don't know what the market is but it must be astronomic. I mean, every room at Vanderbilt has some sort of thing inside it now so we must consume, you know, tanker trucks every day of this stuff. Tom?

DR. FLEMING: There is actually quite a lot I would like to comment on so what I would like to do is just be very brief right now on Frank's comments about the Bayesian methods.

It seems to me that this is an interesting discussion but I am not sure it gets at the essence of what our current challenge is. We are faced

with a mountain of data and, yet, the vast majority of these studies are reported to us in ways that there are significant flaws in the design and conduct--lack of randomization and lack of having vehicles and active controls, and ability to address the many confounding variables, and lack of standardization of product use, and lack of proper handwashing, and surrogates that based on the evidence that we have here don't seem to be correlated with clinical outcomes. I wish the solution to that was statistical, that there would be a magical statistical method that we could use.

A frequentist approach basically says in the context of the data that we have from a given trial, what is the strength of that evidence to establish benefit. We have confidence intervals and values that lie outside that confidence interval or values that are inconsistent with the data. As Frank says, gee, that could be useful in interpreting the study in terms of its strength of evidence but how do we aggregate data?

A Bayesian will come up with their

judgment of other evidence and form a prior and use the data in the trial to form a posterior. That is a very useful approach to look at aggregating evidence. A frequentist also has useful approaches for aggregating evidence using meta-analyses, but does want to keep the purity of the strength of evidence of each individual registrational trial and then allow each of us to use our own subjectivity in how we aggregate the data.

My concern is that my prior could be very different from yours and, hence, my posterior is very different from yours and why should you be committed to my posterior if you don't believe in my prior?

So, in essence, it is an interesting statistical debate and, yet, the essence of our challenge here isn't going to be solved by that debate. The essence of our challenge is do we have integrity in the evidence that is put before us, and how do we aggregate that evidence? And Bayesian methods or frequentist methods can be helpful here but neither is going to get us out of

the morass that we have at this particular point in time due to the lack of having high quality studies that give us the kinds of insights that we would need to answer the questions.

DR. WOOD: Right. Ralph?

DR. D'AGOSTINO: Maybe I am hoping to say what you were going to say, I am hoping to leave by about 5:00--

DR. WOOD: That is exactly what I was going to say! Let's move directly to the questions. I will read the first question to you. Please discuss the use of surrogate markers for the assessment of the effectiveness of healthcare antiseptics.

I guess we should add to that, or maybe implicit in that is the use or not of clinical endpoints within these things, it would seem to me. Is that your feeling?

DR. SNODGRASS: Yes. I have a comment about that, which is how far away is the surrogate marker from the endpoint you are really concerned about? So, yes, you need some sort of clinical

endpoint. I think one of the analogies brought up earlier--I can't remember the specifics but they were saying that surrogate markers have been used but my take on that was that that we are so far removed here--this log count is quite far removed from infection transmission. When you are transmitting from a hand, or whatever, to a patient there is such a gap there for that surrogate marker that that is part of what we have been struggling with for so long.

I guess my comment about this question to assess the effectiveness, well, if the surrogate marker is so far away from the actual clinical goal here, then it can't be nearly as effective and I think that is what we have been struggling with, and that is why it gets back to you need a clinical trial or some type with an endpoint that is of some obvious clinical relevance.

DR. WOOD: Any other comments on question one? Ralph--remembering what you just said!

DR. D'AGOSTINO: Exactly, and I will be sharp and crisp. Just following up on that, we

don't have any evidence that the surrogate leads to clinical endpoints. We just don't have it.

DR. WOOD: Dr. Leggett?

DR. LEGGETT: My take on how we came up with the 1, 2 or 3 logs is because that is what we did with antibiotics when we first noted that they could kill bugs in the test tube. We started off with 10 3 bugs; we killed them all, and that is how we get to 10 3 as bactericidal. I wasn't around then but I can see that that is how we made the leap to saying if we kill 3 logs in the test tubes and we kill 3 logs on the hand we are doing better. So, I think there is some logic. It is not totally false so at least there is a little bit of rationale.

In the development of these sorts of things, I think it would behoove industry if they could show proof of concept in an animal model. It would sort of lend a lot more credence to the fact that that might work in people since infections in animals presumably come the same way as people. And, you could kill a lot of mice without

disturbing an IRB.

DR. WOOD: Right. Tom?

DR. FLEMING: Actually, I apologize in advance, I have a somewhat lengthy answer to this but it sets up the entirety of what I want to say so if I could jump in--

DR. WOOD: Right.

DR. FLEMING: The answer is structured as what is it that makes the surrogate here complicated; what would do we know based on the current data about the reliability of the surrogate; what do we do know from a regulatory perspective; where do we want to be in the future; and what do we need to do to be where we want to be in the future? So, essentially, I think all these are parts to question one.

Quickly, as I think about the factors that could influence how the microbiological effects, the biomarker effects, might impact infection risk, and these are things that I find are critical to think about if you want to look at a biomarker as being predictive of an effect on a clinical

endpoint there, is the degree of effect and that is what we are banking on. Everybody is saying can we use the level, the log reduction as the essence of what is capturing how an intervention is going to be affecting the clinical endpoint? It is plausible that that is one component, but is 10

7

dropping 10 5 the same as 105 dropping 103?

Secondly, the durability of effect is important. We want fast acting; we want persistence. Those are different elements. The breadth of effect matters. Is it broad spectrum? How are we affecting gram-positives? How are we affecting gram-negatives? And position, on the fingernails; in the crevices or deep below superficial skin levels--all of these are complications to this.

There is also the artificial testing conditions that we have in the way we go about trying to assess log reductions. The vigor of scrubbing impacts what log reduction you are going to get. The use of the neutralizers and are we doing that in a consistent way influences?

We are using Serratia instead of what actually might be the bugs that are causing the problems, which are staph. and strep., which can

therefore lead to potentially underestimating and overestimating. Maybe we are underestimating because the effect on Serratia is less than staph. and strep. Conversely, we may be overestimating.

There could be numerous other factors. You might be creating opportunistic influences as you are altering one organism and creating an opportunity for excess growth of another organism that could have a different virology. There is just a wide array of these different types of factors that actually, when you think about this in the totality, doesn't make it too surprising that when the FDA has done their 1,000-article overview what they are finding is not very good evidence that reductions in microbial counts are predictive of effects on infection.

So, the evidence that we would have would suggest that the multidimensional aspect of all of this indicates that what we really care about,

which is a treatment effect on preventing infection, may readily not be reliably addressed by the simplicity of the log reduction since the actual antimicrobial effect that you could have could be much more complex than just summarized in that simplicity.

So, where does that leave us? My own sense is, to answer this question directly, taking a measured strategy, I would think maintaining the current standard for those products that are currently under review is a measured step. But I would hope that we would put into place studies that allow us to have much better insight in the future, insight that is going to allow us to avoid unnecessary healthcare cost if soap and water, together with sophisticated ancillary care, is enough or, if it isn't enough, to recognize what it is that really will provide additional benefit.

Michelle Pearson pointed out that it is possible to do trials that will allow us to look at how interventions affect outcome. She referred to numerous studies, studies on perioperative oxygen,

glucose control, optimal time shaving, systemic antibiotics. We were able, as she was indicating, to do properly controlled trials to be able to understand how these factors influence infection risk. It certainly ought to be possible, therefore, to do such studies to be able to find out whether or not these antibacterial agents affect risk.

So, I would throw on the table some proposed strategies that I think could be feasible, and obviously would need to be fleshed out between statisticians at the agency and industry. But I would argue that designs to look at efficacy or effectiveness could be very useful. An efficacy comparison would be, for example, handwash, a randomization where everyone has handwash and there is a blinded assessment of the vehicle against the antibacterial intervention. This would be a superiority trial that would be blinded.

On the other hand, an effectiveness study that would be an open-label study looking at the antibacterial against an active control, such as

handwashing, would also be a very important trial and it could be done as a superiority trial.

As Dr. Fischler pointed out, in the healthcare personnel handwash setting the unit of randomization would be the hospital unit, and one could be randomizing surgical intensive care units, and you would need about 50-100 of these where you would be looking within each unit about 50 patients or so. So, we are looking at trial sizes that are much like the Parienti trial that we were looking at.

In this context, it sounds daunting but these are large, simple trials. These are trials where you don't take each of these participants and go through the intensive antimicrobial assessments. You are looking at outcomes that are basically is there an infection or is there not where you would take a random sub-sample of these participants, but only a small fraction, and do the antibacterial assessments so that you can carry out the kinds of analyses that Dr. Powers was talking about, that is, within these trials, what is the effect of the

intervention both on infection rate as well as on the biomarker?

In the patient preop skin preparation, a very similar approach could be taken where now the patient is the unit of randomization, so that becomes simpler, and it could be an open-label trial because you could now have a blinded evaluator who is separate from the caregiver, the person who is administrating the intervention.

It has been indicated that the surgical hand scrub situation is the most controversial of these as to whether we could do it, but I would put on the table the possibility of randomizing to soap and water with vehicle versus the antibacterial in a blinded trial as a study that, from what I have heard, I believe could, in fact, still be an ethical trial.

The question then is who is going to pay for these studies? Who is going to do these trials? Well, in fact, who has done the studies that have been adequately powered to look at infection endpoints? Certainly, the hope would be

that there would be a combination of industry support for these trials together with government and NIH support.

Within the last couple of weeks I was asked to testify before the Senate as to what might be done to allow the FDA to be more effective, and one of the things that I suggested was to provide FDA funding for a program that would enable the FDA to ensure that there are observational and clinical trial studies done where these funds in particular could be useful to conduct important studies that would be controlled trials for widely used products, the setting that we are in right now, where there isn't, in fact, the assurance that they are going to be done in a timely way by industry and NIH. So, I would argue this is one such setting.

The bottom line is what I would hope we would do is identify what is correct and what ought to be done, and advocate for what ought to be done and hope that that advocacy for what ought to be done will motivate those people that do have the

potential to do the right thing to, in fact, pursue that.

DR. WOOD: Good. I guess all of these trials that were done in surgical settings were done with all the complexities that exist for every other one and, in fact, it was possible to demonstrate the things that altered the effect, including time of administration, which is normally a difficult demonstration to make in a trial. So, it is possible to do these trials. I agree. Mike?

DR. ALFANO: Just with a clarification because CDC promulgated the guidelines that this group is suggesting has an unacceptable database. So, I don't think we should have the presumption that the study she was talking about, about controlling diabetes for example or sugar levels and the like, would necessarily pass mustard for this type of review. So, we just need to be careful because all the studies we talked about were published, for the most part, in peer-reviewed journals. The studies she talked about were published in peer-reviewed journals. We just don't

know that her studies would pass mustard under this type of review.

DR. WOOD: Other comments? Yes, John?

DR. POWERS: I can assure you that the systemic antibiotics that are approved for perioperative prophylaxis did pass our mustard and are approved for exactly that. So, shaving and things like that--I don't think FDA approves, you know, razors but at least for the systemic antimicrobial drugs, those were exactly the same data that we used to approve those for those indications.

DR. WOOD: Dr. Larson?

DR. LARSON: I just want to point out one other design issue that is slightly different. Actually, I think the studies that you are suggesting in OR are much easier than studies on clinical units. The difference is the intervention. You give an antibiotic; you know you gave it; you know the dose; you watch and you can watch every time it is done. You shave; you know you shaved or didn't shave, or whatever; you know

it is done.

When you are doing a hand hygiene intervention on a clinical unit and you have 70 different people who touch every patient every day, you have to make sure that everybody who comes onto that unit follows the protocol to which they are assigned. That is the problem. That is the problem because you have, as you saw, per nurse 43 indications, or per ICU, 43 indications for hand hygiene, whatever it was that Dr. Boyce showed, per hour and you have to make sure 24 hours a day that everybody who is assigned to one thing does it. That is the difference in intervention. It is a little bit more complicated but I agree with you that it can be done and we have done one, as I said, which is going to be coming out in Archives very soon, and more can be done.

But even the Parienti paper which, in my opinion, is the best one and the only clinical trial that has ever been done in surgery was just dissed here because, well, it was comparing alcohol and CHG and, you know, maybe if we can convince

somebody to do a plain soap that would be, I guess, the answer.

DR. WOOD: But you wouldn't necessarily start on the ward unit; you would start in places where you could do your studies most easily and if you demonstrated an effect in that setting you would move down to other--

DR. LARSON: And where would that be where you have a clinical endpoint?

DR. WOOD: Well, surgical scrubs for a start.

DR. LARSON: Oh, well, he was just saying surgical would be the hardest. I am saying it is not. Surgical products and surgical studies are a little bit different than handwashing or hand hygiene studies clinically. That is where things are used a lot. My question is, we are talking now about OTC products--at least they are right now, where there is no opportunity for industry to patent anything. So, why would they spend money for a clinical trial?

DR. WOOD: What do you mean?

DR. LARSON: Unless they are under an NDA.

DR. WOOD: Right, if they are under an NDA, which is what we are talking about--

DR. LARSON: Oh, this is OTC setting.

DR. WOOD: If they come in--wait a minute, guys, before you all laugh. If you come in with an application that shows that you reduce bacteremia and bring that in under an NDA you can patent that.

DR. LARSON: Under an NDA, but we are talking about OTC products now, how you look at endpoints for OTC products, unless we want to change those to not be OTC.

DR. WOOD: Well, we are encouraging you to do both. Tom?

DR. FLEMING: Yes, just to clarify, when you are looking at the patient perioperative skin preparation we are agreeing. I am saying the simplicity of that is that the patient is the unit of randomization. When you look at the surgical hand scrub setting, I am not claiming this is difficult in terms of unit of randomization. There I would have the surgeon as my unit of

randomization. What I was claiming was difficult were comments that some have made as to whether they would accept soap and water as an appropriate control regimen. If that is appropriate, and I am putting it on the table that I am not persuaded that we have enough evidence to say it can't be, then I think this would be a very viable study where you would look at soap and water vehicle versus soap and water with the antibacterial in a blinded trial.

You are right. In the healthcare personnel handwash what I was indicating was I would randomize by the hospitalization unit for the very reasons you are talking about, and we would, in fact, encourage that entire unit to use the strategy that we are comparing. If that strategy is, in fact, looking at something based on an active control such as handwashing versus an antibacterial, my own view of that is I want to educate and work with that group to achieve a high level of real-world adherence but it doesn't have to be 100 adherence because I am looking at

effectiveness. I want to know the answer, what is the relative effectiveness of a strategy based on the antibacterial where I am educating and encouraging in that unit--

DR. LARSON: Ah, but now you have added the intervention of education and now you have a multifactorial intervention. I mean, this is exactly what we are saying the problems with the studies are.

DR. FLEMING: But I don't view it as a problem at all. I view this as the real-world aspect of what I want to know the answer to. If I implement a strategy within a unit that is advocating the use of this antibacterial versus an active comparator control, this is the answer I want; it is the exact thing we do in many settings. In our HIV/AIDS prevention trials it is the same thing where you can say there is a behavioral component. That is inherently part of the story. I want that factored into the design.

DR. LARSON: But that was a criticism of many of these studies.

DR. FLEMING: The criticism, for example of the Parienti trial, was that it was looking at two different interventions.

DR. LARSON: Not the Parienti trial but a lot of the others were criticized because of multiple interventions at the same time, like education and just those things you are talking about.

DR. FLEMING: Well, it depends on the manner in which that is incorporated and the manner in which they are controlled. If it is a properly randomized, controlled trial looking at effectiveness, then it is not a criticism.

DR. WOOD: Which most of them weren't. Most of them were serial trials.

DR. FLEMING: That is right, and then it becomes a much different issue.

DR. PATTERSON: Some of them weren't but that was still the criticism.

DR. WOOD: Any other comments on question one?

[No response]

I guess we don't need to vote on that so let's move on to question two, has compelling evidence been provided to change the currently used threshold log reduction standard? Please vote on each product category separately.

Okay, has compelling evidence been

provided to change the currently used threshold log reduction standard? Anyone want to start on that? Ralph?

DR. D'AGOSTINO: I don't see any evidence--again, back to the surrogate, we don't have any way of tying in the particular endpoints with effectiveness. So, I don't see how we have any way of sort of pulling back from what is already in the monograph.

One of the things that I do have difficulty with, and it is because I am caught up with not following the logic, is in the healthcare personnel handwash products, the wash 1, 2, 3 4, up to 10. I just haven't heard anything that says that that is compelling one way or the other in terms of keeping it or dropping it. I just would

like to hear what other people have to say about that. But, anyway to summarize, I don't see anything that the sponsors have said that would say that we have evidence that we should change and drop the level of requirement, and I do have this other comment about the multiple washing. I just didn't hear enough in terms of what we are getting at by having it.

DR. WOOD: Dr. Leggett?

DR. LEGGETT: My thought about the 10 washes is that people are going to wash their hands 10 times. If it is only 10 times a day, it is still 10 washes. So, I want to make sure that we don't do damage to the efficacy/safety part of that so I would like to keep those 10 washes in there to make sure that on the 10th one the hands aren't so cracked that it is worse. Conversely, I don't understand why it has to be 3 out of 10 instead of just 2 out of 10.

DR. WOOD: Mary?

DR. TINETTI: Actually, I was going to say something very similar to Dr. Leggett. I think the

advantage of the multiple wash--we are hearing that they should be washing 40 times a day so if they wash 10 it would be nice to know that there is actually an increase that, at least theoretically, could be extrapolated to the number of washes that they should do. Again, whether it needs to be higher than the first wash, but I think seeing the multiple washes does extrapolate to some of the clinical issues.

DR. WOOD: Dr. Larson?

DR. LARSON: We have cultured--I don't know, 8,000 nurses' hands over periods of years, etc. The average count now on nurses' hands--granted, there tend to be more women and smaller hands so the counts are a little smaller because of the square surface area, but the average counts are 4-5 logs when they come to work. If you are expecting a 3-log reduction you are not going to get it. You are starting at such a low number now that I am not sure you are going to be able to see it, and I don't see any rationale for having a need for increased reduction after 10. You want

the hands to be as clean as they can be every time you touch a patient from the beginning wash and there is no reason, that I can see, why it should be better after 10 washes.

DR. PATTERSON: Regarding the specific question about has there been compelling evidence provided to change the currently used log reduction standard, I think the answer to that is no.

But I do think there is a compelling argument or case to evaluate it for change based on the fact that in the TFM the standards are set arbitrarily and are not evidence-based. I would favor looking at persistence. I don't think that cumulative needs to be looked at for efficacy but should be looked at for tolerability and safety. Getting back to the issue again of the clinical trials, I think that would be ideal. As far as the handwash and preoperative skin preparation, if our federal agencies can advise the accrediting agencies that accredit us that we don't need to monitor handwashing or antisepsis, then perhaps that will be feasible.

As far as the surgical hands scrub, based on 20 years of infection control and the infection control literature that has numerous reports of

outbreaks, particularly in the OR, that have been linked to flora found on the hands and shown to be the same organism, I think that there is good enough data to say that it would not be ethical in a developed country where antisepsis is available to have a trial that used a vehicle instead of an antiseptic.

DR. WOOD: Dr. Bradley?

DR. BRADLEY: It seems as though voting on this monograph is going back to what the FDA said--the monograph was designed to deal with drugs which were on the market before the '70's. If we vote to keep this current monograph, which is probably not relevant to new studies coming forward, how much of these criteria in the current monograph will be applied for new drug applications? So, in a sense, if we vote for this and industry doesn't want to do something along these lines, would they go through an NDA process

which would be more strict than this or more flexible, and it would be like redesigning the monograph from scratch but not through this process?

DR. JOHNSON: This gets to be the chicken and the egg problem. We have been told by our general counsel that, in looking forward, if we finalize the monograph we could in similar scenarios have to apply the same criteria to NDAs, that is, until we got to the questions you posed before about significant changes in the products, significant changes of the indication, and then we would bring forward different criteria.

Let me just clarify, when I am referring to pre-'72 it is active ingredients on the market pre-'72. Products using those active ingredients can come forward under the monograph as new products. They are not NDAs but they are new to the marketplace; they just use the same active ingredients. A product that had a completely new active ingredient would have to come in under an NDA and could most likely use these criteria.

Again, it goes back to your earlier questions about how different it is and what indication they are seeking, and that sort of thing, but if they are trying to toe the same basic line, same criteria.

DR. WOOD: Mike?

DR. ALFANO: Yes, I am just troubled by slide number 11 that Dr. Fischler showed which was that chlorhexidine did not, at least in his trial, pass the current TFM. So, if that were finalized--admittedly that wouldn't be involved because it is an NDA product, but presumably everything else that wasn't NDA would go away. Is that true? If that is true, how comfortable are we if that monograph is to be finalized?

DR. WOOD: Why doesn't the FDA respond directly to that question?

DR. LUMPKINS: Because the monograph is finalized doesn't mean that all the products go away. Obviously, you have NDA products out there that can continue to market. Also, products can be reformulated to comply with the monograph standards. So, it is a question of reformulation,

maybe even relabeling.

DR. ALFANO: A follow-up to that, I think the problem is that the newest version of the monograph includes a cleansing wash. To Dr. Larson's point, that wash reduces the burden to the extent that there was no log reduction in the first wash with 4 percent chlorhexidine product. So, that troubles me if, in fact, that is the way it is to be applied. Now, it could be changed as it goes to final monograph. If you take the wash out maybe that is a different scenario.

DR. LUMPKINS: Exactly. The monograph methodology is not engraved in stone. There are a lot of issues that we heard today about this methodology and we are certainly going to try and rectify a lot of that if we continue to go down this road. So, we are aware of the problem with that extra handwash in the handwash methodology and it is totally unvalidated.

DR. WOOD: And there were lots of other problems that were raised--

DR. LUMPKINS: Yes.

DR. WOOD: There were lots of other problems raised with the actual methodology that would need to be addressed. That is not a question

that is here and I don't think its absence should imply that the committee is endorsing the methodology.

DR. JOHNSON: Just with regard to the personnel handwashes, just to clarify, the original wash is to take away some of the factors associated with the actual physical properties of the skin such as oiliness and that sort of thing. Also, the personnel handwash methodology involves the inoculation. So, the mentality is that you are kind of getting everyone to a cleanliness state, whatever that might be, and then inoculating them to a similar higher level. At least, that is the theoretical basis for it.

DR. WOOD: Any other comments? Frank?

DR. DAVIDOFF: I have a general comment. I think it applies more to the personnel handwash than to the two surgically related ones. This strikes me as very much like a lot of clinical

decisions where there are harms and benefits to either side of the decision. I mean, if the standard is relaxed, it seems to me that wouldn't preclude someone from coming up tomorrow with a new agent that actually was more effective and, in fact, would meet whatever standard we thought was good. But a relaxed standard still would allow the development of better agents, if that is one of the general goals. Someone could also figure out how to get 100 percent compliance with the existing agents which would probably do quite a bit to reduce clinical infection.

On the other hand, if the relaxed standard were adopted it would remove, I think, some of the incentives to develop better products because you don't have to beat such a tough standard. Not relaxing the standard, keeping it as rigorous as this, seems to me would keep only the most "effective" agents on the market and it might force the search for better agents.

On the other hand, it might, as has been discussed, remove a lot of agents that really

probably are doing something useful, which would be really a fairly major concern. Another part of the downside is that if the standard were maintained as very strict, the people in the industry might very well see that that is a standard that is going to be hard to meet and they might just simply leave the industry altogether because the likelihood of, you know, putting in money to develop the product that met the standard might simply be seen as not feasible.

So, I am struggling not so much on the basis of the science but on the basis of the implications, the potential benefits and harms, particularly in the absence of the clinical infection data.

DR. WOOD: Dr. Leggett?

DR. LEGGETT: I thought we were only still talking about personnel handwash but I will just jump in for the other two.

DR. WOOD: Let's do them all at once.

DR. LEGGETT: Okay. My comments about not doing wash 2 and wash 11 are the same that I had

for wash 10 in the personnel handwash. I don't understand--my same point--why it has to be 3 logs at wash 11 5 days later. What is the logic? Does that mean that eventually somebody is going to have sterile hands at a month and a half? I mean, that is not going to happen.

The other thing I had is about sticking a needle through somebody's chest. How is that different pathogenetically than putting a scalpel to their stomach? So, I don't understand why we need 2 logs in the stomach but only 1 log if we are going to put a big hemodialysis catheter in their chest.

Then, I am not sure why we need 3 in the groin, except that there are more bugs there so it is easier. However, if we want to look at any clinical surrogate endpoints, we know that there are no more line infections from groin lines than there are from subclavian lines. So, how can that square?

Given all that, if the CFU decline doesn't mean anything, and there is not a lot of good data,

I don't see any reason to change it, in other words to decrease it.

DR. WOOD: Mary?

DR. TINETTI: We have been hearing all day that there is no relationship between these log reductions and the outcomes that we are interested in--

DR. WOOD: I think you needed to be on another planet not to get that information from this. Tom?

DR. FLEMING: Well, reading the question literally, for me it is an easy answer, is there compelling evidence to change the currently used log threshold, no, no and no. Now, the issue, is going beyond that, what do we think about this--

DR. WOOD: Well, let's deal with just the question first because we have to vote on it, that is why. Well, go ahead.

DR. FLEMING: Well, briefly and it is an issue that has been stated before, it has been correctly noted by a number of colleagues around the table, all right, but we don't really have

compelling evidence to say why it has to be a larger level of protection when you have additional washes. Of course, I also don't know whether 1 or 2 is enough. And, my general sense in working with surrogates is that I have a great deal of concern about their use unless there is the level of reliability of validation that we have talked about, but my intuition says when in doubt, the larger the level of effect you are asking for, it does influence plausibility that you are actually going to get protection.

So, in the serious absence of evidence here, if we are still going to be using these measures, it strikes me as illogical to be weakening what it is when we are saying that what has been put forward itself hasn't been justified. My sense as well is if, in fact, what we are putting forward is a standard that is rigorous, might that rigorous standard provide indirect motivation for people to do the kinds of trials we really want? We have made it very easy for three decades based on a relatively weak standard for

people to not enter into the kinds of trials that will really reliably tell us what types of interventions and what types of biological effects truly will provide patient protection. So, it seems to me this wouldn't be the time to weaken a standard when we have acknowledged that this standard itself hasn't been rigorously justified.

DR. WOOD: So, picking up on Mary's comment and on yours, would it be the committee's pleasure to have a question of has compelling evidence been provided to justify the current standard? Is that what you want? And then take that second question? Or do you just want to go to that question? Is that what you are saying?

DR. LARSON: Could I just ask--of course, I am not voting, but I just want to ask the committee why you think there haven't been studies done. It seems to me that one compelling reason to ask is this, if this has been the standard since 1978 why have the studies not bee done?

DR. WOOD: Let me answer that. I can reel them off and I can keep us here all night, but

studies were not done comparing diuretics to standards in antihypertensive therapy. There were no studies done comparing a placebo to postmenopausal estrogens. There are lots of studies that were not done and there were all kinds of reasons for why they were not done. It does not necessarily mean they are impossible to be done.

DR. LARSON: No, of course not, but it might mean that the surrogates are not very meaningful to the people who are getting the money to do the studies.

DR. WOOD: Right, I agree, and that is what I think Tom and I are saying, that we here to motivate them to get it done.

Hearing no compelling evidence that we want two votes, let's take one. Has compelling evidence been provided to change the currently used threshold log reduction standard? The answer to that would be that if you wanted to keep the standard you would say no, and if you wanted to change the standard you would say yes. Agreed?

DR. FLEMING: Not quite. I mean the

question doesn't say that. The question just says has compelling evidence been provided.

DR. WOOD: Right.

DR. FLEMING: That is all it is saying.

DR. WOOD: All right. So, has compelling evidence been provided to change the currently used threshold log reduction standard? We will go down A, B and C. To make it efficient, let's do them in one round so we don't have to go around three times. Let's start with Dr. Leggett.

DR. LEGGETT: By A you mean handwash?

DR. WOOD: Yes, sorry. Handwash would be A; the surgical scrub would be B, and the patient preoperative skin preparation would be C.

DR. LEGGETT: So no one forgets that we are trying to herd cats, I will say A, no; B, no; C, no. But I would like FDA to consider some tweaks, as I mentioned.

DR. D'AGOSTINO: No on all three.

DR. TINETTI: No on all.

DR. BLASCHKE: No on all.

DR. WOOD: Dr. Larson is not voting?

DR. LARSON: I am a consultant.

DR. LUMPKINS: You can vote. You have voting privileges.

DR. LARSON: No, except maybe for the cumulative issue. That is a subset of two of them.

DR. WOOD: All right. Wayne?

DR. SNODGRASS: No on all three.

DR. PATTEN: No on all three.

DR. WOOD: No on all three.

DR. PATTERSON: No on all three, except for the cumulative data.

DR. BRADLEY: No on all three except the day 5 surgical scrub.

DR. CLYBURN: No on all three, except the cumulative.

DR. FINCHAM: As the questions are listed, no on all three.

DR. FLEMING: No on all three.

DR. DAVIDOFF: No on all three.

DR. WOOD: Let's go on to question number three, given the current standards using surrogate markers to demonstrate efficacy, how should the

analysis be conducted?

How should we define meeting the threshold, for example mean log reduction, median log reduction, percentage of subjects meeting threshold?

How should we evaluate the variability in the data? And, how do we evaluate the variability in the test method?

These are long questions. Anyone want to start off with that? Yes, Ralph?

DR. D'AGOSTINO: I realize the present TFM is ambiguous and we probably aren't going to straighten things out completely, but in terms of the type of endpoints and designs within the log reduction that I think makes sense, if we make a suggestion they do a mean log reduction, I think that is fine.

I think that also percent subjects meeting threshold has a lot of merit to it and certainly a lot of clinical trials run two primaries or one primary and an important secondary. So, I think both of those as endpoints make a lot of sense.

As far as variability of the data, I think that we should suggest and what I think should be done is that we start looking at confidence

intervals of these values, not just that you attain a mean. When you talk about variability of the test method, there are a lot of different ways of handling it but one design that was mentioned by the presentation of the FDA was to have a vehicle and an active control plus the test so you have a three-arm study. I am not sure I follow completely what it means to have a vehicle here, if that is possible or what-have-you, but I think that that type of design, a three-arm study with a vehicle--some type of low-level activity; what does the vehicle actually do; what does soap and water actually do as one arm. Another with the active control, and then the test.

And then the study in terms of the analysis to handle the variability of the test method you would look at the active versus the vehicle; you would look at the test versus the vehicle and that would be a way of getting the

internal validation of the study. You want the active to work in this study. In addition to that, we would want both the active and the test to exceed the bacterial reduction criteria or percent criteria, whichever we felt was appropriate, the most important endpoint. So, it is looking at a three-arm study, getting internal validation and then also getting some real comfort and solid support that you have also maintained the bacterial reduction.

DR. WOOD: Let's take each question separately. Let's do meeting the threshold question first. Any further discussion on that that people have? Tom?

DR. FLEMING: Well, just sticking to that answer, it is certainly very appropriate I would say to advocate for any one of these three approaches. The two that seem most appealing to me are the ones that we probably use the most, which is the mean log reduction, but then also looking at the percent meeting the threshold has a real appeal to it. I think Dr. Valappil did a very nice job of

laying out these pros and cons.

The concern with the mean log reduction is that it is possible that you could have some outliers that create a favorable mean. Let's say you wanted a 3-log reduction, you might be achieving that but heavily influenced by a few outliers. So, the alternative of looking at the percent of subjects that meet the threshold is very appealing if, in fact, we have a pretty good sense that what you really need for protection is--I will throw out a number--a 3-log reduction. Anything less than that isn't protective; anything greater is. Then, clearly, in that scenario I would want to look at the percent of subjects that meet the threshold.

In the absence of really having a good sense about this, the disadvantage of that is you are throwing away some information that the mean is keeping. So, my own sense is I could advocate for either of those two approaches because they have relative merits.

DR. WOOD: Dr. Leggett?

DR. LEGGETT: If you kept the mean but then you included confidence intervals, that would solve the problem that was presented by the FDA,

wouldn't it?

DR. FLEMING: I am going to jump ahead and strongly agree with Ralph that the confidence interval is critical here. So, it is a very important feature but it doesn't necessarily get around the influence of outliers that you would still have when you are looking at means.

DR. FINCHAM: Alastair, aren't we making assumptions about measures of central tendency of percentages of individuals meeting a threshold without any consideration of sample size? If you have a sample size of two, none of these are going to be effective, in my mind. So, I don't know if that clouds the issue more but appropriate statistical techniques and research design, in my mind, mandate that you have appropriate sample sizes.

DR. WOOD: Right. Presumably, you would have to have some power calculation to determine

the difference that you were going to be able to exclude. So, I think inherent in this is the assumption that we are going to have some predefined power calculation that says what sort of difference we are going to be powered to exclude. I would think that but I will defer to Ralph and Tom.

DR. D'AGOSTINO: Yes, the reason I was answering all three is because I do agree that you have to respond to all three in order to think what the study is going to be like. When you get down to the third one, if you are talking about a vehicle you are saying the active versus the vehicle must be statistically significant and so you must have big enough sample sizes for the test. I agree a hundred percent with what you are saying.

DR. WOOD: Tom?

DR. FLEMING: Again I agree with Ralph that for me, as a statistician, the answer to parts one, two and three is an integrated answer. Just to reiterate, the answer to part (i) is very difficult in the absence of believing in this as a

marker that we really adequately understand as to how it is predicting benefit.

The answer to (ii), as Ralph has already said--it seems to me that point estimate, as important as it is, is our best sense of what the data tell us about the effect. The precision of that estimate is critical. You have to understand not just the point estimate but the precision and, hence, the confidence interval becomes really key.

The third aspect of this is how do we evaluate the variability in the test method? My own sense about this is I think there is more than one way that you adequately do this so I want to kind of quickly walk through three steps. One way to do this is to compare the test against a vehicle. This would typically be in a setting where it is a blinded trial and you are wanting to look at efficacy. Clearly, in that setting I want superiority, and I would want superiority at the level that the guidelines have indicated. But as industry has mentioned, therefore, what is the lower limit of the confidence interval that you

would accept? At this point I would consider, in the spirit of what has been stated, that the lower limit of the confidence interval has to rule out this 20 percent lesser effect than the target effect, which more or less is going to mean your point estimate is going to have to be close to the target effect or better to rule that out.

A second design would be looking at the test against an active comparator. That could be either an open-label effectiveness trial or a blinded efficacy trial. The ideal here would be superiority again. The ideal would be if there is superiority and I can show superiority, then I am comfortable having just those two arms. The concern that existed with the Parienti trial is that when there isn't superiority against an active control you don't know whether you are equally effective or equally ineffective. But if you have superiority those data are interpretable.

The third third approach is when you are going head-to-head with the test against the active comparator can it be good enough just to show

non-inferiority? Technically, yes. Technically, it can if i know the active comparator is providing substantial effect and that effect is precisely understood. Then, in fact, I can come up with a margin. But here is the essence, I have to know that there is assay sensitivity here. I have to know, in the context of the trial in which the active comparison is being done, that this active comparator is providing substantial benefit in order to be able to justify a non-inferiority margin.

So, a variation of that design when I can't be that confident would be the three-arm study that people have been talking about. You do the test and the vehicle and the active and the vehicle and essentially that strategy allows, when it is ethical, when it is ethical to have a vehicle--it allows you to be able to look directly at test against vehicle and have the active in there to basically validate assay sensitivity. But, in essence, I need that third arm in a setting where I can't be confident that I know what the

efficacy is of the active comparator.

I would accept as well in this setting that if I know the active comparator is highly effective, then I can use a non-inferiority margin and still be confident that I am establishing effect at the level that is targeted.

DR. WOOD: But in the absence of knowing that you almost would have to have--

DR. FLEMING: In this third strategy, in the absence of being confident that I know that the active comparator is going to be highly effective at a defined level, then I don't have the assurance of having assay sensitivity. That is when I have to insert then the active comparator arm in with the test and vehicle into three arms. DR. WOOD: Mike?

DR. ALFANO: Just something that may help people formulate their perspectives, you know, chlorhexidine is cationic so it is formulated with cationic surfactants which are not as good at cleansing as are the anionics and, therefore, when you look at the vehicle control chlorhexidine has a

bit of a built in advantage versus its own control.

DR. WOOD: So, what you are saying is you need the appropriate control, whatever that is. I don't think Tom was implying that it was necessarily the vehicle.

DR. ALFANO: A straight vehicle control would make it look better. I am not knocking chlorhexidine, mind you, but it is a technology issue.

DR. WOOD: Got it.

DR. ALFANO: The other comment, thinking back to my days in microbiology, for problems of this type you want to keep a large number of products available, presumably products that work, of course. If you look at the data we have reviewed you have seen scenarios presented where people were having trouble on the ward when they were using chlorhexidine. When they switched to alcohol it improved. When they were using alcohol and switched to chlorhexidine it improved. So, we just need to be careful that we don't lose those abilities to switch as problems arise given an

endemic infection in a given hospital setting.

DR. WOOD: Don't you think when these switches were made, you know, multiple interventions occurred simultaneously?

DR. ALFANO: Well, that is the criticism--

DR. WOOD: When there is an outbreak like that everybody suddenly wakes up and says, wow, we had better do what we are supposed to be doing.

DR. ALFANO: It could be.

DR. WOOD: Right. Frank?

DR. DAVIDOFF: Yes, first I just should mention that Tom is clearly a Bayesian because he keeps saying how confident he is.

But, no, I had a specific question for the agency to follow-up on Tom's point that there is valuable information both in the mean and in the percent of subjects meeting the threshold. My question is whether it is considered appropriate and useful, or even possible, to use a dual criterion in some fashion, that is, both measures or some combination of those measures rather than just one or the other. I can see how it might

create difficulties to create a rule that you have to either meet both or, if you don't meet both, one of them has to be above--I mean, it could get more complicated. On the other hand, not using both might lose important information.

DR. POWERS: It is possible. The issue is if you are going to have two endpoints and apply equal weight to those, that usually entails some adjustment of what your test of significance is to be able to do that.

But the question that we struggle with is, is the information that we are losing significant information in terms of what Tom said. We don't know that we need to really differentiate the person who has a 6-log reduction from a 5-log reduction. I guess that is what we struggle with. The percent of subjects achieving a threshold really kind of addresses the mean piece because you will be picking up that information. As Thamban said, it won't allow us to differentiate the people who have huge reductions from less huge reductions. The question is, is the information that is lost

there worth knowing and, unfortunately, we don't have the answer to that. So, I guess what we struggle with from a clinical perspective is we are worried that we may have the example Thamban showed where you have 4/18 people who actually achieved the mean log reduction driving the entire results. In that case you lose even more important information, in that the vast majority of the people there did not achieve what you wanted in terms of that surrogate.

DR. WOOD: Any other comments on this question? I think we have worked that to the end; we don't need to vote on that. So, question four, the last question, current labeling for healthcare antiseptics consists of class labeling that does not include product performance information. What labeling information would be helpful for clinicians to fully understand product efficacy?

Well, from my perspective one that would clearly be important for clinicians would be to demonstrate that it actually produces some clinical effect. So, that would be the highest hierarchial

point for me and I would see that as being of such a different standard that it would get an NDA approval and would potentially have huge commercial and public health advantages. I can't see any reason not to tell people how well it does in the surrogate either. I think it was Tom who made the point earlier that that drives people to perform better. What do other people think?

DR. CLYBURN: Having read this, I calculated that as I was seeing patients yesterday, I washed my hands 40-some odd times and I was using an alcohol wash and I didn't feel terribly confident, having read all of this, that there was a lot of data to support what I was doing. I think I would like to know that. I might choose something else.

DR. WOOD: Right. Yes, John?

DR. POWERS: One of the things we wanted to address here that we weren't able to capture in the question was exactly what you mentioned, should we differentiate between products that say they have met a specific threshold in terms of a

surrogate, but this has not been demonstrated to be proven to decrease infections in a clinical trial from products who actually go out and do that?

DR. WOOD: I think they are different products. The others would become--no pun intended--some soap that you could buy over-the-counter. It would be hard to imagine a hospital buying that product if there were ones out that had a demonstrated hard endpoint. Yes, Dr. Leggett?

DR. LEGGETT: A question for the FDA again, so this would not be the sort of thing where a product, say, triclosan named A did better than triclosan named B. In other words in the current monograph it would be based on this log reduction. So, say, triclosan company A goes out and they get 2.8 logs and company B gets 2.3--

DR. POWERS: That is not what we were suggesting. Since we don't know the clinical impact of that, if you met the crieria--

DR. WOOD: Just like everybody else did.

DR. LEGGETT: Because you would be

inundated by all sorts of people--

DR. POWERS: Right, as opposed to saying you met the criteria and you actually demonstrated a clinical benefit.

DR. WOOD: I think if you have demonstrated clinical benefit the issue of meeting the criteria is irrelevant, frankly. I don't think these are linked. I didn't mean them to sound linked. Tom?

DR. FLEMING: John, I don't know if I am going further than what you are saying. What I had written down here was I would like to reward those sponsors that have taken the high road and have done the rigorous studies to provide more conclusive assessments about efficacy as well as activity. So, shouldn't the label say something to the effect that this intervention has achieved the targeted 3-log reduction in X percent of patients and healthy volunteers relative to control, but clinical studies have not established whether there is a decrease in infection rate? So, specifically indicate what has been established and what hasn't

been established. Then, when another sponsor comes along and has established, it is very clear and part of the reward for the effort to go through the process of identifying not just the effect on biomarkers but on clinical efficacy endpoints is that their label clearly reflects that distinction.

DR. WOOD: Absolutely. Other comments? If not, then at 4:48 we are adjourned.

[Whereupon, at 4:48 p.m., the proceedings were adjourned.]

- - -